A Linear Model of Wintertime Low-Frequency Variability. Part I: Formulation and Forecast Skill

Christopher R. Winkler, Matthew Newman,
and Prashant D. Sardeshmukh

NOAA-CIRES Climate Diagnostics Center, University of Colorado, Boulder, Colorado

ABSTRACT

A linear inverse model (LIM) suitable for studies of atmospheric extratropical variability on longer than weekly time scales is constructed using observations of the past 30 years. Notably, it includes tropical diabatic heating as an evolving model variable rather than as a forcing, and also includes, in effect, the feedback of the extratropical weather systems on the more slowly varying circulation. Both of these features are shown to be important contributors to the model's realism.

Forecast skill is an important test of any model's usefulness as a diagnostic tool. The LIM is better at forecasting Week 2 anomalies than a dynamical model based on the linearized baroclinic equations of motion (with many more than the LIM's 37 degrees of freedom) that is forced with *observed* (as opposed to the LIM's predicted) tropical heating throughout the forecast. Indeed at Week 2 the LIM's skill is competitive with that of the global nonlinear medium-range forecast (MRF) model with nominally $O(10^6)$ degrees of freedom in use at the National Centers for Environmental Prediction (NCEP). Importantly, this encouraging model performance is not limited to years of El Niño or La Niña episodes. This suggests that accurate prediction of tropical diabatic heating, rather than of tropical sea surface temperatures *per se*, is key to enhancing extratropical predictability.

The LIM assumes that the dynamics of extratropical low-frequency variability are linear, stable, and stochastically forced. The approximate validity of these assumptions is demonstrated through several tests. A potentially limiting aspect of such a stable linear model with decaying eigenmodes concerns its ability to predict anomaly growth. It is nevertheless found, through a singular vector analysis of the model's propagator, that predictable anomaly growth can and does occur in this dynamical system through constructive modal interference. Examination of the dominant growing singular vectors further confirms the importance of tropical heating anomalies associated with El Niño/La Niña as well as Madden-Julian oscillation episodes in the predictable dynamics of the extratropical circulation. The relative contribution of initial streamfunction and heating perturbations to the development of amplifying anomalies is similarly examined. This analysis suggests that without inclusion of the effects of tropical heating, extratropical weekly averages may be predictable about two weeks ahead, but with tropical heating included, they may be predictable as far as seven weeks ahead.

1. Introduction

The fact that the patterns of variation of weekly and longer-term averages are distinct from those of daily weather suggests that the mechanisms of their generation are also different, and perhaps predictable beyond the synoptic predictability limit. Several such mechanisms have been proposed and studied over the past several decades, such as zonal-eddy index cycles, resonant interactions with topography, regime-like behavior, synoptic-eddy feedbacks, modified planetary-wave dynamics in the presence of a spatially varying ambient flow, response to slowly varying tropical diabatic heating, and slowly varying boundary conditions at the ocean, sea-ice, and land surface boundaries. Most of these studies

Corresponding author address: Matthew Newman, NOAA-CIRES Climate Diagnostics Center, Mail Code R/CDC, 325 Broadway, Boulder CO 80305-3328, USA.
E-mail: matt@cdc.noaa.gov

have been of a diagnostic nature, concerned with pointing to the particular mechanism under consideration as an essential ingredient of a more complete theory of low-frequency variability, rather than with constructing such a complete theory. They have thus failed to provide a comprehensive framework in which to compare theory with observations, and to assess the relative importance of the mechanisms listed above. They have also stopped short of providing closed models for *predicting* the low-frequency variations. Such closed models are not only needed for basic understanding and prediction but also to guide the development of general circulation models.

Simmons et al. (1983) made a notable first attempt at constructing such a model, by focusing on barotropic Rossby wave dynamics as modified in the presence of a meridionally and zonally varying background 300 hPa flow. In their calculations, many of the low-frequency perturbation eigenmodes of the linearized 300 hPa barotropic vorticity equation (i.e. Rossby waves) had spatial structures broadly resembling those of observed atmospheric teleconnection patterns, and several were unstable and thus likely to be naturally selected. In view of this, Simmons et al. suggested a trivial closure assumption, that the forcing of these Rossby waves could be treated as random. Borges and Sardeshmukh (1995), however, questioned the instability of the waves in the presence of boundary layer damping and therefore the natural selection process. Sardeshmukh et al. (1997) found that the linear and nonlinear barotropic models had only modest forecast skill. Newman et al. (1997) put the linear model to a less severe test of predicting just the zero-lag and time-lag circulation statistics, and also found it to be wanting. The primary limitation was the neglect of the spatial and temporal structure of the forcing. The forcing required to produce realistic statistics (as shown in Figure 3 of Newman et al.)

has relatively large amplitude in the Rossby wave source regions associated with tropical diabatic heating (Sardeshmukh and Hoskins 1988) and in regions of large synoptic-eddy feedbacks. The forcing is clearly also not white in time, thereby begging the question of what causes *its* coherent variation. In other words, to close the model it is necessary to model the evolution of some aspects of the forcing, instead of specifying it all as incoherent white noise.

The modeling of tropical diabatic heating is feasible on seasonal and longer scales because of its close association with tropical sea surface temperatures (SSTs), which are themselves predictable several months ahead without explicit knowledge of other system variables (e.g. Penland and Sardeshmukh 1995). On shorter scales, however, this association breaks down. Heating can exhibit strong variability even in the presence of steady anomalous SST (e.g. Sardeshmukh and Hoskins 1985, Zhang and Hendon 1997, Sardeshmukh et al. 2000). Also, the dominant mode of tropical intraseasonal variability, the Madden-Julian Oscillation (MJO), is not driven primarily by SST anomalies. The modeling of extratropical synoptic-eddy feedbacks presents similar difficulties. Storm-track models now exist for predicting synoptic-eddy statistics and feedbacks associated with a slowly varying ambient flow (e.g., Branstator 1995, Whitaker and Sardeshmukh 1998). As discussed by Whitaker and Sardeshmukh, however, on scales of an individual season and shorter this association too becomes weaker.

One reason for the reduced interest in simple theoretical modeling relative to global numerical simulation has been closure difficulties such as above. We believe, however, that progress can be made by relaxing the requirement of strict closure and allowing for errors that may be treated as temporally (but not nec-

essarily spatially) incoherent white noise. If the statistical moments of the noise are state-independent ("additive" noise), they may be treated as external constants; no further closure is necessary. If they are state-dependent ("multiplicative" noise), then that dependence must also be modeled for closure. In practice considerable progress can be made by assuming state-independence of the noise. In the context of modeling the synoptic eddy feedbacks above, this amounts to assuming that the *error* in parameterizing the feedback, not the feedback itself, is independent of the slowly evolving circulation. Note also that explicit retention of noise in the feedback immediately implies predictability limits, consistent with Whitaker and Sardeshmukh's analysis. Similar remarks apply to the modeling of tropical diabatic heating and other forcing.

One could thus embark on a program of modeling extratropical low-frequency variability, with this extended interpretation of closure, through a hierarchy of simplified, stochastically forced dynamical models. In this paper we have constructed one such stochastically forced linear model indirectly using the *observed* zero-lag and time-lag statistics of 7-day running-mean Northern Hemisphere circulation and tropical diabatic heating anomalies. Such an inverse modeling procedure (Penland and Sardeshmukh 1995; Penland 1996) yields a Linear Inverse Model (LIM) with a dynamical operator that accounts for all the linearly parameterizable dynamics of 7-day running mean anomalies. Even processes not explicitly considered are implicit in the dynamical operator to the extent that they are *parameterizable* in terms of the inverse model's variables. Thus, LIM attempts to determine the same operator that potentially could be obtained, in a forward sense, from the nonlinear dynamical equations. In fact, to the extent that nonlinear processes are parameterized, the model provides a better linear representation of a dynamical system

than a linearized version of that system's governing equations (Branstator and Haupt 1998; DelSole and Hou 1999). By construction the inverse method models low-frequency flow as a stable, linear system driven by geographically coherent white noise. These assumptions are testable, however, and in general we find them to be valid. Note that although the dynamics are asymptotically stable in this model, disturbances can grow over finite time intervals either through the interference of nonorthogonal eigenmodes or in response to the stochastic forcing.

Linear inverse models have proven useful in several geophysical applications, most notably in studying ENSO (Penland and Magorian 1993; Penland and Matrosova 1994; Penland and Sardeshmukh 1995; Johnson et al. 2000) and tropical Atlantic SST variability (Penland and Matrosova 1998). Several investigators have also constructed linear inverse models from atmospheric data, though much of the focus has been on identifying principal oscillation patterns (e.g., Xu and von Storch 1990; von Storch and Xu 1990; von Storch and Baumhefner 1991). Penland and Ghil (1993) found modest improvement in short-term predictive skill relative to persistence forecasts in a LIM of lowpass-filtered, annually-varying Northern Hemisphere 700 hPa geopotential height anomalies. However, including all seasons may have degraded their results, since a pronounced seasonality exists in the atmospheric response to remote forcing (Opsteegh and Van den Dool 1980; Van den Dool 1983; Newman and Sardeshmukh 1998). The evolution of 700 mb heights clearly also depends upon processes not directly parameterizable in terms of 700 hPa heights alone, such as tropical convection and upper tropospheric dynamics. We improve upon Penland and Ghil's results here by focusing only on wintertime variability, by considering both barotropic and baroclinic components of the flow, and by

incorporating tropical diabatic heating as an evolving model variable. Heating is included explicitly since many previous studies (e.g., Lau and Phillips 1986; Sardeshmukh and Hoskins 1988; Ferranti et al. 1990; Kiladis and Weickmann 1992; Qin and Robinson 1995; Higgins and Mo 1997; Mo and Higgins 1998; Mo 1999) have demonstrated that tropical forcing has an impact on extratropical low-frequency variability.

A central question of this work is whether a good *linear* model of low frequency variability is feasible. We find that the answer is yes. The success of our approach is evident in Fig. 1. Details of the calculations are given in section 4, but one need not know them to grasp the main point. The panels display forecast skill in Week 2 of the following linear models: unforced barotropic, unforced baroclinic, tropically forced baroclinic, unforced LIM, and tropically forced LIM. The skill of persistence forecasts is also shown. All forecasts are initialized with an observed 7-day mean anomaly, and then run for 14 days. Thus 2280 forecasts (76 initial conditions x 30 winters) are summarized in each panel. What is displayed is the anomaly correlation at each gridpoint between the forecasts and their corresponding verifications, the observed 7-day running mean anomaly at Day 14.

It is immediately clear that the LIM is much better than models in which anomalies grow and develop only through barotropic and/or baroclinic interactions with the base state. In fact, the unforced model forecasts are less skillful than persistence. Including tropical diabatic heating as forcing greatly improves the linear model forecasts (cf. Fig. 1d with 1c and Fig. 1f with 1e), but significant improvement can also be made by including linearly parameterized internal processes such as transient eddy feedback (cf. Fig. 1e with 1c and Fig. 1f with 1d).

These comparisons are necessary because forecast skill is an important test of any model's usefulness as a diagnostic tool. Given this evidence that the LIM is a good forecast model (indeed, its skill will be shown to be comparable to that of a state-of-the-art GCM), it provides a means for finding dynamically relevant structures in atmospheric data. Also, by treating diabatic heating explicitly, one can distinguish between those structures more dependent on tropical diabatic heating and those more dependent upon other processes. In order to carry out such an analysis one must, of course, first formulate the LIM and test its ability to make good forecasts. This is done in sections 2 and 3. In section 4 the forecast skill of the LIM is compared with that of a nonlinear medium-range forecast model in use at the National Centers for Environmental Prediction (NCEP). We find that LIM forecasts are reliable at the medium range, though potential predictability exists for much longer lead times. Tests of the validity of the LIM, considered in section 5, further support our hypothesis that large-scale extratropical variability during northern winter may be viewed as stable linear dynamics forced with white noise. The impact of tropical heating upon extratropical low-frequency variability is assessed in section 6. The *predictable* growth of extratropical circulation anomalies is shown to depend crucially upon tropical diabatic heating. Without the heating, long-term anomaly growth is not predictable, and short-term growth is of weak amplitude. A discussion and summary are presented in section 7.

2. Linear inverse modeling

Consider an atmospheric state vector $\mathbf{X}$. We define anomalies as $\mathbf{x} = \mathbf{X} - \bar{\mathbf{x}}$, where $\bar{\mathbf{x}}$ is some base state, typically a time mean. Then the evolution of $\mathbf{x}$ may be represented as

Anomaly correlation of 14 day forecasts
Based on forecasts made for DJF 69/70-98/99

(a) Barotropic model

(b) Persistence

(c) Baroclinic model

(d) Baroclinic model, tropically forced

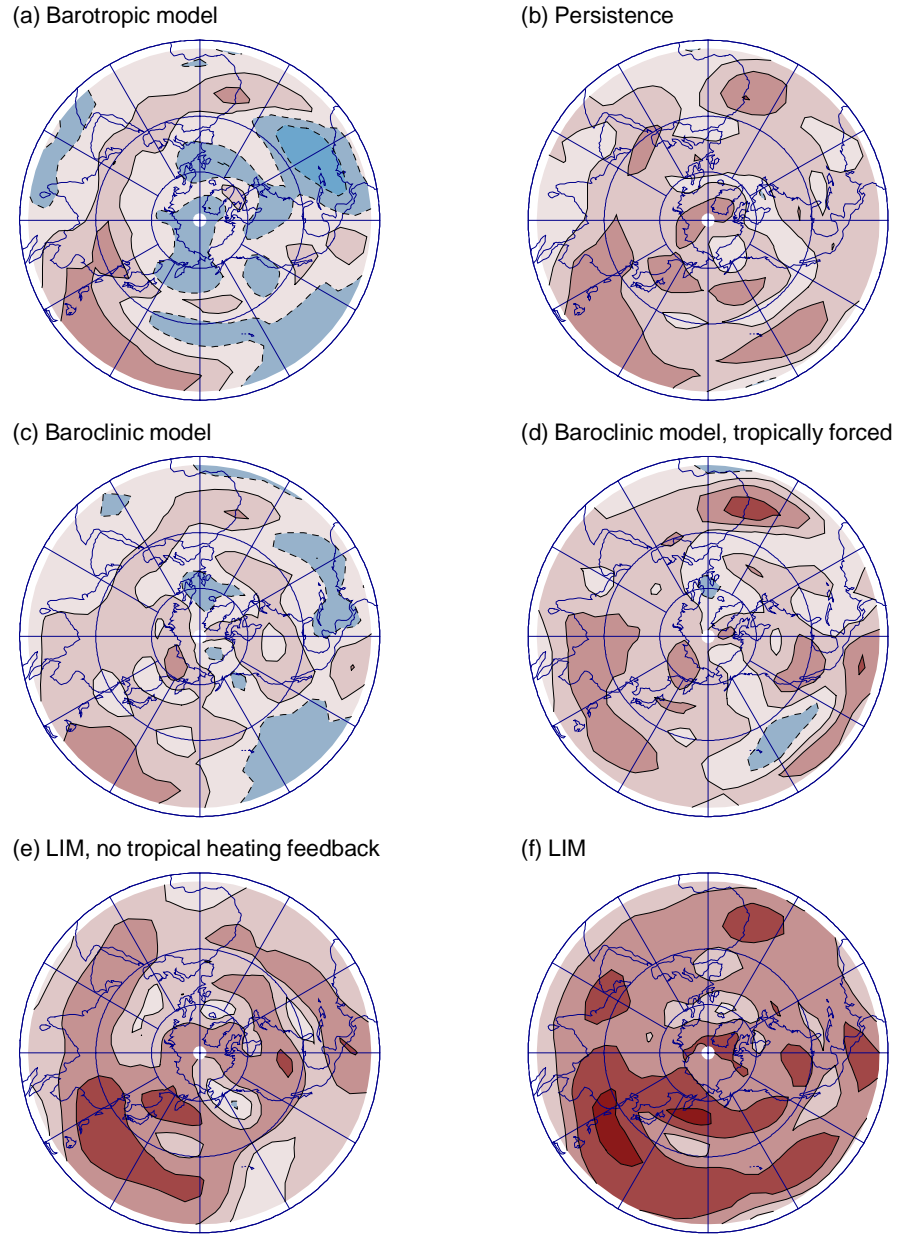(e) LIM, no tropical heating feedback

(f) LIM

Figure 1: Anomaly correlation of week 2 250 hPa streamfunction forecasts produced from the same initializations for the winters 1969/70-1998/99. (a) Unforced barotropic model. (b) Persistence. (c) Unforced baroclinic model. (d) Tropically forced baroclinic model. (e) Linear inverse model with the effects of heating upon the evolution of stream-function eliminated. See section 6 for further details. (f) Linear inverse model. Contour interval is 0.15 with negative and zero contours indicated by dashed lines and blue shading. Red shading denotes positive values of correlation, with the reddest shading indicating values above 0.6. In all polar stereographic figures shown in this paper, the outermost latitude is the equator.

$$\frac{d\mathbf{x}}{dt} = \mathbf{L}\mathbf{x} + \mathbf{N}(\mathbf{x}) \qquad (1)$$

where **L** is the linearized part of the dynamical equations and $\mathbf{N}(\mathbf{x})$ represents nonlinear terms.

Note that $\mathbf{L}$ and $\mathbf{N}(\mathbf{x})$ depend upon $\bar{\mathbf{x}}$. In general, the statistics of $\mathbf{x}$ may not be understood without a detailed knowledge of $\mathbf{N}(\mathbf{x})$. In some highly nonlinear systems, however, for suitable temporal and/or spatial averaging $\mathbf{N}(\mathbf{x})$ may be approximated as $\mathbf{N}(\mathbf{x}) \approx \mathbf{T}\mathbf{x} + \mathbf{F}_s$, where $\mathbf{T}$ is a linear operator and $\mathbf{F}_s$ is noise which is white in time but which may be spatially correlated. Thus under this averaging (1) may be expressed as

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \mathbf{F}_s \qquad (2)$$

(e.g., Papanicolaou and Kohler 1974; Hasselmann 1976; Penland 1996). In this paper, $\mathbf{x}$ represents low-frequency eddies, $\mathbf{T}\mathbf{x}$ represents the linearly parameterizable nonlinearity (including synoptic eddy feedback), and $\mathbf{B} = \mathbf{L} + \mathbf{T}$. Note that in general, $\mathbf{T}$ need not be diagonal.

The solution to (2) is
$$\mathbf{x}(t + \tau) = \mathbf{G}(\tau)\mathbf{x}(t) + \sigma(t + \tau), \qquad (3)$$

where τ is some lag after initial time t,

$$\mathbf{G}(\tau) = \exp(\mathbf{B}\tau), \quad \text{and} \qquad (3a)$$

$$\sigma(t + \tau) = \mathbf{G}(\tau)\int_t^{t+\tau} \mathbf{G}(t')^{-1}\mathbf{F}_s(t')dt'. \quad (3b)$$

Since $\sigma(t+\tau)$ does not depend upon $\mathbf{x}(t)$, the best forecast $\widehat{\mathbf{x}}(t + \tau)$ (in the least squares sense) given $\mathbf{x}(t)$ is

$$\widehat{\mathbf{x}}(t + \tau) = \mathbf{G}(\tau)\mathbf{x}(t). \qquad (3')$$

This is also the most probable state if $\mathbf{F}_s$ is Gaussian (e.g., Penland 1989). In either case, $\sigma(t+\tau)$ is the forecast error.

Now $\mathbf{B}$ could, in principle, be determined from an appropriate linearization of the dynamical equations plus a dynamically-based linear parameterization of the nonlinear terms. Alternatively, $\mathbf{B}$ may be indirectly estimated from the observed simultaneous and lagged statistics of $\mathbf{x}$ at some lag τ_0. Since the forecast error is uncorrelated with the initial state, it can be shown using (3) that

$$\mathbf{B} = \tau_0^{-1}\ln\{\mathbf{G}(\tau_0)\} \qquad (4)$$

$$= \tau_0^{-1}\ln\{\mathbf{C}(\tau_0)\mathbf{C}(0)^{-1}\}$$

where

$$\mathbf{C}(\tau_0) = \langle \mathbf{x}(t + \tau_0)\mathbf{x}(t)^{\mathrm{T}}\rangle$$
$$\mathbf{C}(0) = \langle \mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}}\rangle \qquad (5)$$

and angle brackets denote an ensemble average. Although τ_0 is the lag used to calculate $\mathbf{B}$, forecasts may be made for any lag τ using (3'). If $\mathbf{x}$ has stationary statistics then by construction $\mathbf{B}$ is a stable operator (Penland and Sardeshmukh 1995). That is, all eigenvalues of $\mathbf{B}$ have negative real parts and hence all eigenmodes of $\mathbf{B}$ eventually decay.

Nevertheless, deterministic eddy growth can occur (and does occur, as we will see in section 6) over finite time intervals through nonmodal interaction (e.g., Farrell 1988). Since $\mathbf{B}$ is stable, however, the eddies modeled by (2) must decay *on average* (see Sardeshmukh et al. 1997). Such dissipation must be balanced by energy input from stochastic forcing if $\mathbf{x}$ is to have stationary statistics. This balance is represented by a fluctuation-dissipation relationship derived from (2),

$$\mathbf{B}\mathbf{C}(0) + \mathbf{C}(0)\mathbf{B}^{\mathrm{T}} + \mathbf{Q} = 0 \qquad (6)$$

(e.g., Gardiner 1985; Penland and Matrosova 1994). The matrix $\mathbf{Q} = \langle \mathbf{F}_s\mathbf{F}_s^{\mathrm{T}}\rangle dt$ is the covariance of stochastic forcing multiplied by dt. An important application of (6) is to deter-

mine $\mathbf{Q}$ given $\mathbf{B}$ and $\mathbf{C}(0)$. Eigenanalysis of $\mathbf{Q}$ then reveals the principal patterns of stochastic forcing, sometimes referred to as noise EOFs. We defer detailed discussion of $\mathbf{Q}$ to the second part of this study (Newman et al. 2001b). Here it is only important to note that since $\mathbf{Q}$ is a covariance matrix it must be positive definite; that is, all its eigenvalues must be real and positive. This is an important test of LIM (Penland and Sardeshmukh 1995).

The technique used to determine $\mathbf{G}$ is unremarkable, in that it is simply multiple linear regression (MLR) applied at some lag τ_0 in (3), which is then used to determine $\mathbf{B}$ in (4). The primary goal of LIM is to determine whether (2) is a good model of the observations, and if so to determine dynamically meaningful system structures through an analysis of $\mathbf{B}$. If one were interested only in making forecasts, then $\mathbf{G}(\tau_0)$ determined by MLR for various τ_0 would suffice. What distinguishes LIM from MLR is its claim that the underlying dynamics are of the form (2); that is, the $\mathbf{G}(\tau_0)$ for different τ_0 are *related* as $\mathbf{G}(\tau_0)=\exp(\mathbf{B}\tau_0)$ where $\mathbf{B}$ is a *constant* operator. It is thus critical to assess the extent to which $\mathbf{B}$ determined from (4) is independent of the τ_0 used to estimate it, since only then can the underlying dynamical equation be represented by (2). Equations (2)—(6) thus relate the simultaneous and time-lagged *statistics* of $\mathbf{x}$ to the *dynamics* $\mathbf{B}$ that govern it, which is a notable advantage over other empirical techniques such as Singular Value Decomposition or Canonical Correlation Analysis (CCA). These techniques, while useful for statistical forecasting, have considerable difficulty identifying dynamical relationships in geophysical systems (e.g., Newman and Sardeshmukh 1995; Cherry 1996).

In practice, an appropriate dynamical subsystem of the full climate system must be chosen to focus on the phenomena and time scales of interest. That is, we select only certain variables in a truncated space, which we hope to show *a posteriori* encompass most of the variability on the time scales of interest here. In this paper we define $\mathbf{x}$ as

$$\mathbf{x} \equiv \begin{bmatrix} \psi \\ \mathbf{H} \end{bmatrix} \qquad (7)$$

where ψ is anomalous streamfunction and $\mathbf{H}$ is anomalous tropical diabatic heating. Equation (2) in this truncated system can be written as

$$\frac{d}{dt}\begin{bmatrix} \psi \\ \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{\psi\psi} & \mathbf{B}_{\psi\mathbf{H}} \\ \mathbf{B}_{\mathbf{H}\psi} & \mathbf{B}_{\mathbf{HH}} \end{bmatrix}\begin{bmatrix} \psi \\ \mathbf{H} \end{bmatrix} + \begin{bmatrix} \mathbf{F}_{s_\psi} \\ \mathbf{F}_{s_\mathbf{H}} \end{bmatrix} \qquad (8)$$

and forecasts are made from

$$\widehat{\mathbf{x}}(t+\tau) = [\mathbf{G}(\tau_0)]^{\tau/\tau_0}\mathbf{x}(t) \qquad (9)$$

$$= \exp\left(\begin{bmatrix} \mathbf{B}_{\psi\psi} & \mathbf{B}_{\psi\mathbf{H}} \\ \mathbf{B}_{\mathbf{H}\psi} & \mathbf{B}_{\mathbf{HH}} \end{bmatrix}\tau\right)\mathbf{x}(t).$$

Here $\mathbf{F}_{s_\psi}$ and $\mathbf{F}_{s_\mathbf{H}}$ are the white noise forcing of ψ and $\mathbf{H}$, respectively. If $\mathbf{x}$ contained only ψ, (2) might still describe the low-frequency variability of ψ provided $\mathbf{H}$ could be linearly diagnosed in terms of ψ. However, it is only by including $\mathbf{H}$ explicitly in $\mathbf{x}$ that it becomes possible to diagnose how tropical heating impacts extratropical low-frequency variability through $\mathbf{B}_{\psi\mathbf{H}}$ (also see Newman et al. 2000). Including $\mathbf{H}$ also gives better forecasts of ψ, and allows forecasts of $\mathbf{H}$ as well.

Certainly, other variables not included in (7) could also be important to the evolution of ψ or $\mathbf{H}$. The inverse model does, however, implicitly include the effects of all other variables linearly related to ψ and/or $\mathbf{H}$. This is an important distinction with a forward dynamical model in which the evolution of the state

vector is governed only by *explicitly* represented interactions among its components.

3. Model data and analysis technique

Thirty years (1969/70 to 1998/99) of anomalous wintertime (December 1—February 28) streamfunction and diabatic heating data are used to define **x**. The diabatic heating rates were determined from an improved iterative solution of the "chi-problem" (Sardeshmukh 1993; Sardeshmukh et al. 1999). This iterative procedure was applied to twice-daily NCEP Reanalysis wind fields to minimize the nonlinear vorticity budget imbalance at 28 atmospheric levels, and the modified divergent wind circulation was further constrained to satisfy the large-scale mass budget. Diabatic heating rates were finally determined as a balance requirement in the thermodynamic energy budget, using the modified wind circulation to compute the other terms. The LIM calculations were also repeated with the original, uncorrected heating rates and with outgoing longwave radiation data (Liebmann and Smith 1996). Overall, the dynamically consistent heating rates used here gave the best results.

Both ψ and **H** were spectrally truncated to T21 and transformed onto a Gaussian grid. Low-frequency anomalies were defined by removing the first 3 harmonics of each variable's annual cycle at each gridpoint and then applying a 7-day running mean filter. Qualitatively similar results were obtained for 5-day, 11-day, and 15-day running means. Streamfunction anomalies were determined at 750 hPa and 250 hPa. Diabatic heating anomalies were smoothed with a T21 spectral filter which attenuates small-scale features and Gibbs phenomena (Sardeshmukh and Hoskins 1984). Although the vertical structure of tropical diabatic heating has been shown to have some impact on tropical-extratropical interactions (e.g., Ting and Sardeshmukh 1993), we use

column-integrated diabatic heating anomalies from the surface to the tropopause as a single measure of **H** in (7). Examination of the vertical profile of anomalous heating suggests that this simplification is reasonable in areas of deep tropical convection in winter, but is less accurate in regions of large-scale subsidence.

Anomalies were projected onto their leading Empirical Orthogonal Functions (EOFs) to make (2)—(5) computationally tractable while still accounting for much of the variability. Streamfunction EOFs were defined only for the Northern Hemisphere (NH), and diabatic heating EOFs only in the region 30°S—30°N. Prior to computing EOFs, each field was normalized by its domain-averaged climatological root-mean-square amplitude. The EOFs of streamfunction were computed from a vector combining the normalized 750 and 250 hPa streamfunction anomalies, rather than at each level separately.

The leading 30 EOFs of anomalous NH streamfunction and leading 7 EOFs of anomalous tropical diabatic heating were retained. The time-varying coefficients of these EOFs, i.e., the principal components (PCs), define the 37-component state vector **x** in (7). Figure 2 shows that 30 streamfunction EOFs account for over 90% of the low-frequency variability in the extratropics. The 7 heating EOFs capture over 70% of the low-frequency variability in the central and western Tropical Pacific, though only about 35% of the domain-integrated variability. A lag of τ_0=5 d was used to determine **B**. Sensitivity of the results to lag and EOF truncation is discussed in section 5. Note that all the model predictions shown in this paper are compared to *untruncated* data.

Finally, the LIM must be tested on data independent of that used to determine **B**. Estimates of **B** and of forecast skill were cross-validated as follows. We sub-sampled the data record by

a) 750 hPa Streamfunction　　　b) 250 hPa Streamfunction
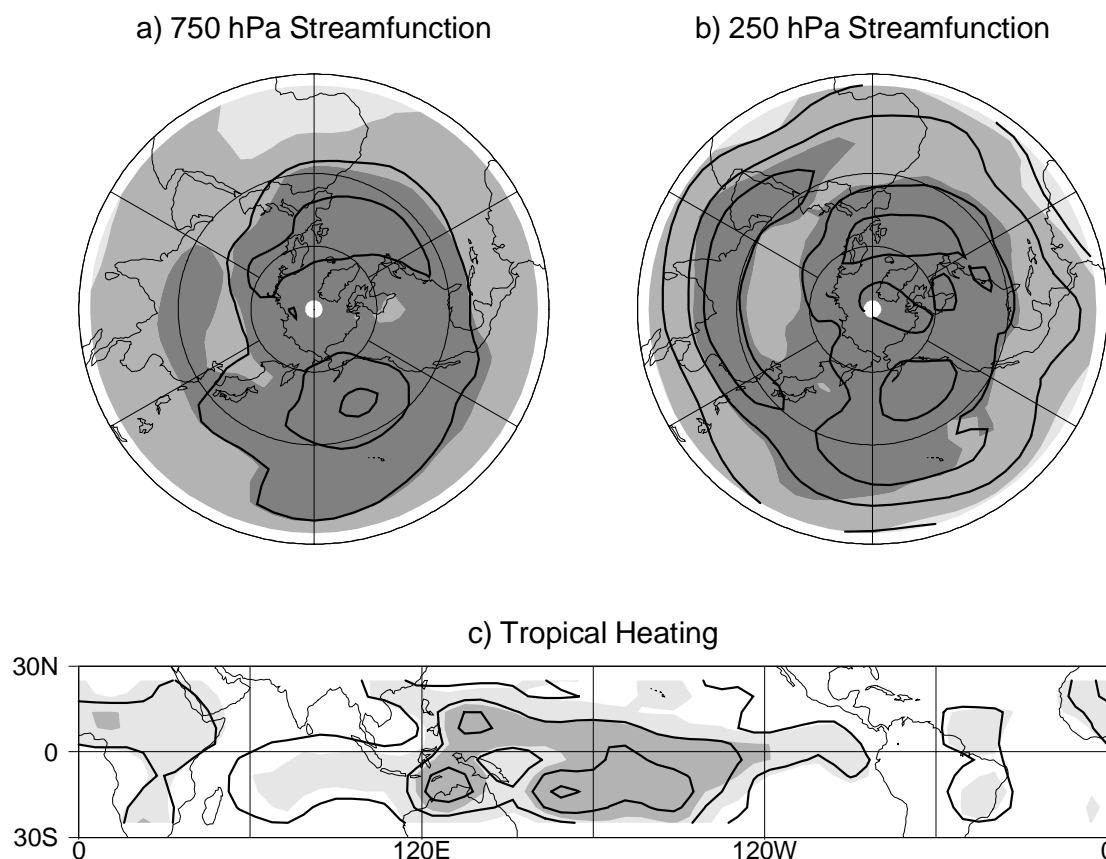
c) Tropical Heating

Figure 2: Variance explained by the EOF truncation used for the linear inverse model. (a) 750 hPa streamfunction, (b) 250 hPa streamfunction, and (c) column-integrated diabatic heating. Light, medium, and dark shadings represent 50%, 70%, and 90% variance explained in each figure. Contours indicate standard deviation of each field from 30 year dataset. Contour interval is 3.5×10^6 m^2 s^{-1} in (a) and (b), 35 W m^{-2} in (c).

removing one of the available years, computed **B** via (4) for the remaining years, and then generated forecasts for the independent year. This procedure was repeated for each year. All measures of forecast skill in this study are based upon these jack-knifed forecasts.

4. Medium-range forecast skill

The most important test of any scientific theory is its ability to predict future observations. That is, *the criterion for a realistic model of low-frequency variability should be its ability to produce good forecasts*. Accordingly, we have compared LIM streamfunction forecasts with forecasts using either climatology, persistence, or the observed local autocorrelation

function. In all cases the LIM displays much greater skill.

A more demanding test is to compare LIM forecast skill with that of numerical dynamical models of increasing complexity. Figure 1 presents such a comparison with persistence and three different models linearized about the same wintertime climatological mean base state: (a) a barotropic model (e.g., Sardeshmukh et al. 1997), (c) a two-level balance (baroclinic) model (e.g., Haltiner and Williams 1980 (pp. 68-9); Whitaker and Sardeshmukh 1998), and (d) the balance model forced with the *observed* evolving untruncated tropical heating anomalies throughout each forecast period. The baroclinic model levels are 250

and 750 hPa, whereas the barotropic model level is 250 hPa. Both models are stabilized by a simplified linear (Rayleigh) damping sufficient to eliminate eigenmode instability. For the barotropic model, the damping time scale is 5 d; but note that the results in Fig. 1a are identical for any damping value, as discussed in Sardeshmukh et al. 1997. We repeated the barotropic model calculation by adding scale-dependent damping associated with the "free-surface correction" (e.g., Ferranti et al. 1990; Qin and van den Dool 1996) but found no notable improvement (not shown). For the baroclinic model, the damping timescale is 20 d in the upper layer and 2.5 d in the lower layer. Weaker damping produced a poorer comparison with observations. Finally, the results of the jack-knifed LIM forecasts are shown in Fig. 1f.

All models were initialized with an observed 7-day mean anomaly, and were then run for 14 days. These forecasts were repeated for each day of the winters 1969/70-1998/99. The displayed measure of forecast skill is local anomaly correlation of 250 hPa streamfunction,

$$\langle \widehat{\psi}(t+\tau)\psi(t+\tau)\rangle /$$

$$\{\langle \widehat{\psi}(t+\tau)\widehat{\psi}(t+\tau)\rangle \langle \psi(t+\tau)\psi(t+\tau)\rangle\}^{1/2}$$

where $\widehat{\psi}$ denotes the predicted 7-day mean anomaly and ψ denotes the verification, for forecast lead τ=14 days.

The LIM has skill above 0.3 everywhere except the North Atlantic. Skill is generally highest over the Pacific basin, with a peak value of 0.64. Clearly, LIM forecasts are everywhere much better than all three linear models. They are also better at 750 hPa, although the pattern of skill is different (not shown). The LIM forecast amplitude is also typically much greater than that of the three linear models (not shown), but the variance of

even the LIM Day 14 forecasts is only about half the observed variance.

Although the LIM has only 37 degrees of freedom, its forecasts are competitive with the ensemble mean forecasts of the NCEP medium-range forecast (MRF) model, a nonlinear GCM with nominally $O(10^6)$ degrees of freedom. Shown in Fig. 3 is a comparison of the Week 2 forecast skill for the winters of
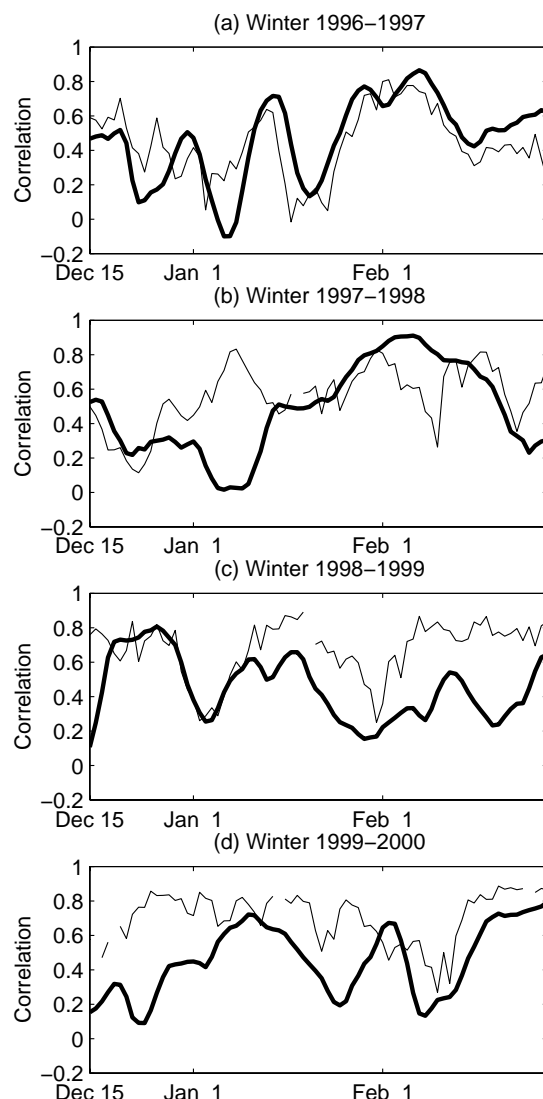

Figure 3: Winter (1996/1997-1999/2000) spatial pattern correlation of Week 2 250 hPa streamfunction forecasts in the region 0--90N, 120E--60W from the linear inverse model (thick line) and the MRF ensemble mean (thin line). Verifications begin 15 Dec of each year.

1996/97-1999/00. The MRF model is the operational version used daily at NCEP during the forecast period. The measure of forecast skill displayed is spatial pattern correlation in the region north of the Equator between 120E and 60W. This allows daily comparison of how well each model predicts the anomaly field in this region. For GCM forecasts, we define Week 2 as the mean of days 8—14 and Week 3 as the mean of days 15—21, where day 0 is the forecast initialization. The LIM forecast is initialized on day -3 with the 7-day mean centered on that date (that is, the mean of days -6 — 0), and is then run forward 14 and 21 days. Thus, the LIM and GCM forecasts verify for the same period, and neither uses information past day 0. As a result, the Week 2 predictions begin verifying on December 15 each winter. Skill of the two models can be quite similar. LIM forecasts are better than MRF forecasts for some extended periods (e.g., January 1997 and February 1998), though also much worse on occasion. Similar levels of LIM skill are seen in the tropics (0°-30°N) and the extratropics (30°N-90°N). Pattern correlations computed over the rest of the Northern Hemisphere (not shown) show a similar level of MRF skill, but LIM skill is reduced about 20%.

The LIM continues to show skill at longer forecast leads, particularly at Week 3. Operational MRF forecasts are not available beyond Week 2, but extended forecasts during the winters DJF 1985/86—1988/89 were made by the Dynamical Extended Range Forecasts (DERF) project (Schemm et al. 1996) using the reanalysis version of the NCEP MRF model. Figure 4 compares Week 2 and Week 3 skill of the LIM and DERF forecasts. Only a single realization (at 200 hPa) of each DERF run was available, so Fig. 4 is not a completely clean comparison. Nevertheless, it is notable that even at Week 2 the LIM skill is higher than DERF skill in the Tropics and comparable, but mostly less, in the extratropics. By Week 3

(Figs. 4c,d) the LIM has higher forecast skill throughout most of the Tropics but now also over much of the extratropics, especially the North Pacific. Similarly, the four-winter average spatial pattern correlation at Week 2 was 0.33 for the LIM and 0.28 for DERF, and by Week 3 it was 0.25 for the LIM and 0.12 for DERF. This suggests that credible forecasts *beyond Week 2* are possible, and are at least partly realized by the LIM.

The LIM has moderate skill in predicting tropical heating anomalies over the 30-year record (not shown). Possibly as a result of the severe EOF truncation, skill is greater in the tropical Pacific than over South America and the tropical Atlantic. Still, the local anomaly correlation of Week 3 LIM heating forecasts with the *untruncated* heating observations reaches as high as 0.7 in the mid-Pacific. Substantial skill is evident in the west Pacific and over most of the MJO region as well. LIM skill is high not only for stationary heating but also for propagating heating anomalies (see Fig. 10). Thus, the LIM heating forecasts can easily beat persistence (not shown) over almost all of the Pacific. However, over Africa the LIM forecasts have substantially lower skill than persistence forecasts (for example, 0.4 vs. 0.6 at Week 2).

Figure 5 shows that LIM skill in the west Pacific is in stark contrast to the DERF skill, which is near-zero in the MJO region (see also Hendon et al. 2000). For the DERF, precipitation is used as a proxy for tropical heating since the forecast heating itself is not available. To facilitate the skill comparison, both the forecast and verification (CMAP; Xie and Arkin 1997) precipitation anomalies were spectrally smoothed in the same manner as the heating. Skill of persistence forecasts during the same 4-winter period is also shown. DERF forecasts of tropical forcing are considerably worse than those of the LIM, with poor skill

Anomaly Correlation of Week 2 Forecasts
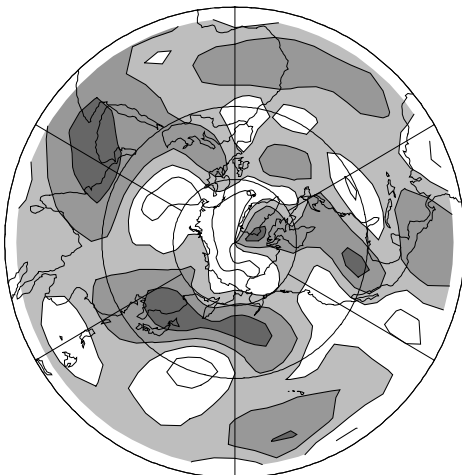
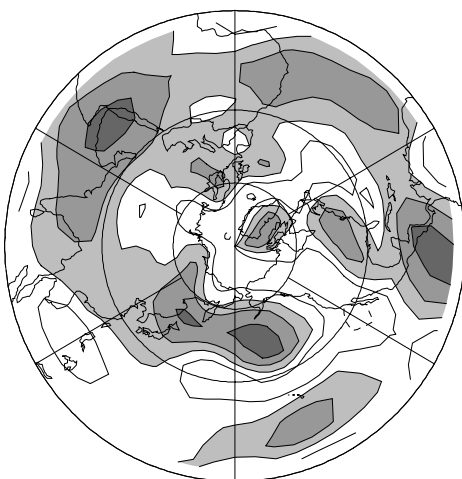### (a) LIM (250 hPa)   (b) DERF (200 hPa)
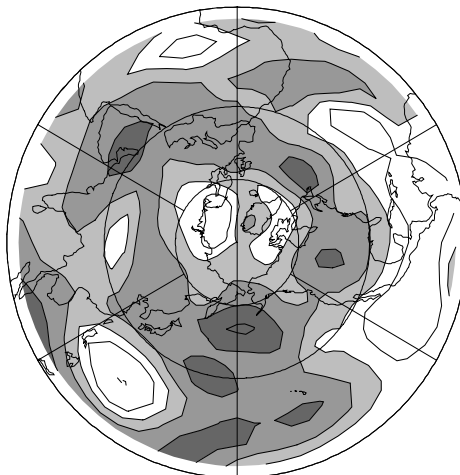
Anomaly correlation of Week 3 Forecasts

(c) LIM (250 hPa)   (d) DERF (200 hPa)
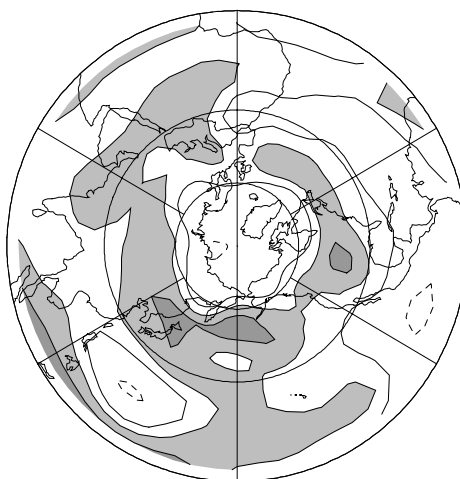
Figure 4: Anomaly correlation of week 2 and week 3 streamfunction forecasts for DJF 1985/86--1988/89. (a) Week 2 LIM. (b) Week 2 DERF. (c) Week 3 LIM. (d) Week 3 DERF. Contour interval is 0.15 with negative contours indicated by dashed lines, and the zero contour omitted for clarity. Values above 0.3 are shaded with darkest shading indicating values above 0.6. Note that the LIM is verified at 250-hPa and DERF is verified at 200-hPa.

evident as early as Week 1. By Week 3 (not shown) LIM skill in the west Pacific is still greater than 0.4, while it is negative for the DERF. The one exception to poor DERF skill is in the equatorial central tropical Pacific, a region where ENSO produces notable convection anomalies.

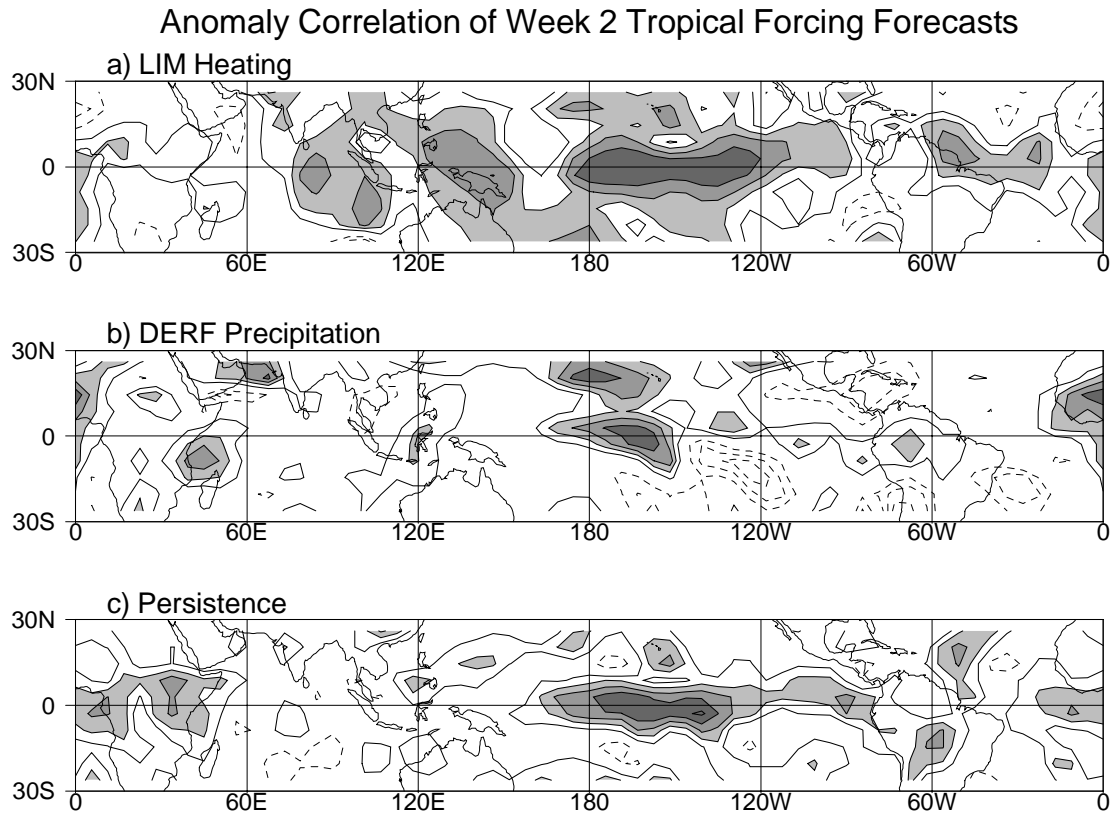## 5. Model robustness

a. Tests for the validity of linear dynamics

Anomaly Correlation of Week 2 Tropical Forcing Forecasts

Figure 5: Anomaly correlation of week 2 tropical forecasts for DJF 1985/86--1988/89. Forecast quantity is column-integrated diabatic heating **H** for the LIM and persistence, and precipitation for the DERF. (a) LIM. (b) DERF. (c) Persistence. Plotting conventions are as in Fig. 4.

As was discussed in section 2, one of the strengths of the LIM technique is that its inherent assumption of linearity can be tested. Forecast skill is only one of these tests. In this section we discuss results from some other tests based upon those in Penland and Sardeshmukh (1995).

1) IS **B** INDEPENDENT OF τ_0 ?

If the system we are modeling were truly stable, linear, and forced only by white noise, then **B** would not depend on the lag τ_0 used in (4). For example, eigenanalyses of $\mathbf{B}(\tau_0)$ and $\mathbf{B}(\tau_0')$, where τ_0 and τ_0' are two different lags, would yield identical results. Also, the forecast $[\mathbf{G}(\tau_0)]^{\tau/\tau_0} \mathbf{x}(t)$ would be the same as $[\mathbf{G}(\tau_0')]^{\tau/\tau_0} \mathbf{x}(t)$. A test which establishes the

independence of the results to the choice of τ_0 is called a Tau-test.

As a practical consideration, however, upper and lower limits of τ_0 exist in the LIM procedure. At very short lag, the time dependence of observed autocovariance cannot be matched by a first-order Markov model, so LIM will tend to underestimate the true decay rates of all eigenmodes (DelSole 2000). On the other hand, as τ_0 is increased important modes of the system are sampled at close to their half period, resulting in aliasing of the corresponding eigenvector of **B** (Penland and Sardeshmukh 1995). Such a lag cannot uniquely determine **B**, although **G** computed at this lag could still be used to make forecasts at this lag; that is, multiple linear regression itself is not invalidated. In other words, consider a system

where (2) is exact. If $2\tau_0$ is a Nyquist period, then given **B** it is always possible to uniquely determine $\mathbf{G}(\tau_0) = \exp(\mathbf{B}\tau_0)$. However, this relation cannot be uniquely inverted to give **B** $= \ln[\mathbf{G}(\tau_0)]/\tau_0$. This Nyquist problem makes it difficult to test the linearity of the system by comparing **B** operators determined using different τ_0, and exists even with an infinitely long data record, since it is due to the time lag τ_0.

Although a Tau-test comparing **B** operators computed from different lags is problematic, a Tau-test comparing **G** computed from different lags is not. For example, the autocovariance

predicted from the LIM (that is, $[\mathbf{G}(5)]^{\tau/5}\mathbf{C}(0)$) can be compared with the autocovariance predicted by multiple linear regression at that lag (that is, $\mathbf{G}(\tau)\mathbf{C}(0)$). In Fig. 6 we show the trace of these predicted autocovariances, separately for streamfunction and for heating, normalized to 1 at zero lag. The correspondence between the two curves in each panel demonstrates that the system is substantially linear.

Likewise, a comparison of 20-day forecasts made using $\mathbf{G}(5)$ (that is, $\mathbf{x}(t+20)=[\mathbf{G}(5)]^{20/5}\mathbf{x}(t)$) to those made using $\mathbf{G}(20)$ reveals little difference, as evidenced by a median global pattern correlation between predictions of 0.8.
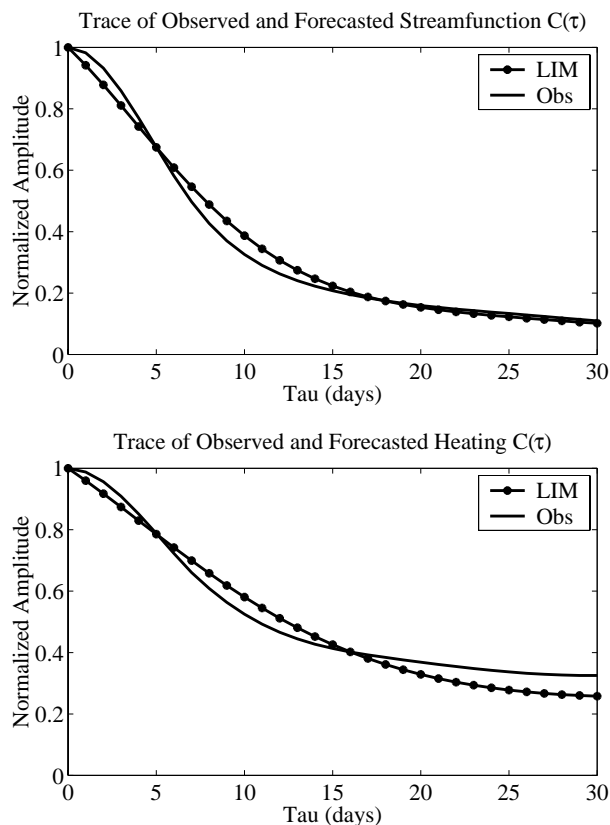

Figure 6: (Top) Tau-test for streamfunction. Thick curve: Trace of $\mathbf{C}(\tau)$ predicted by LIM in principal component space. Asterisks: Trace of observed $\mathbf{C}(\tau)$ in principal component space. (Bottom) Tau-test for heating Thick curve: Trace of $\mathbf{C}(\tau)$ predicted by LIM in principal component space. Asterisks: Trace of observed $\mathbf{C}(\tau)$ in principal component space.
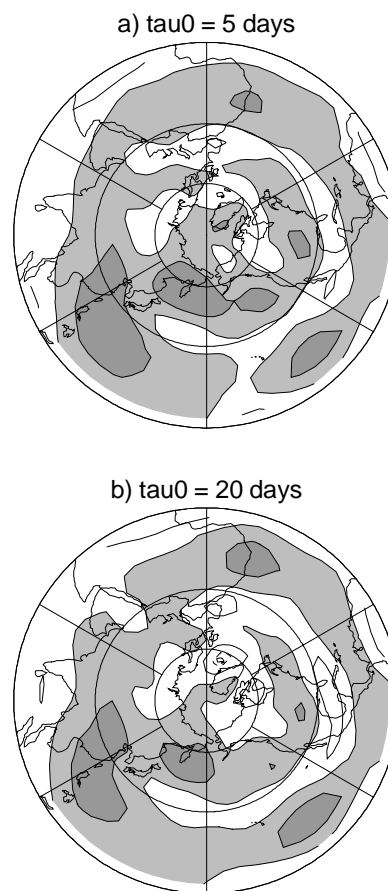

Figure 7: Anomaly correlation of Day 20 forecasts. (a) LIM with τ_0 = 5 d. (b) MLR with τ_0 = 20 d. Plotting conventions are as in Fig. 4.

Indeed, forecast skill (Fig. 7) is slightly better using **G**(5), perhaps because the direct estimation of **G**(20) is more susceptible to sampling errors (see also Johnson et al 2000). This suggests that a LIM, constructed using a relatively short τ_0, can be as skillful as multiple linear regression (or the related CCA technique) at long forecast lead times.

2) DO PREDICTION ERRORS GROW ACCORDING TO THEORY?

Forecasts $\widehat{\mathbf{x}}(t+\tau)$ differ from the true state $\mathbf{x}(t+\tau)$ because of the noise in (2). Nonetheless, if (2) is a useful model of low-frequency variability then forecasts using (3') will be good. Also, the domain integrated forecast error variance, $\delta(\tau) = \langle \sigma^T(\tau)\sigma(\tau)\rangle$, should be the trace of the error covariance matrix, which can be derived from (3) to be

$$\langle \sigma(\tau)\sigma(\tau)^T\rangle = \mathbf{C}(0) - \mathbf{G}(\tau)\mathbf{C}(0)\mathbf{G}^T(\tau) \quad (10)$$

(Penland 1989). In practice, initial condition errors produce discrepancies between $\delta(\tau)$ and the trace of (10) even if **B** truly describes the system's evolution. In the discussion below we normalize $\delta(\tau)$ by the trace of **C**(0) so that the normalized variance $\tilde{\delta}(\tau)$ is bounded by $\tilde{\delta}(0) = 0$ and $\tilde{\delta}(\infty) = 1$.

In Fig. 8, error growth expected from (10) is compared to error growth of the jack-knifed LIM forecasts. Over the first 10 days, LIM prediction error (thick solid line) grows somewhat faster than expected by theory, though the error is still much smaller than both climatology (thin solid line) and persistence forecast errors (open circles). This discrepancy between actual and theoretical error occurs because of a weak correlation between the initial state $\mathbf{x}(0)$ and the forecast error $\sigma(\tau)$ over short lags ($\tau <$ 5 d). If the true noise is somewhat red, rather than strictly white as we have assumed, then
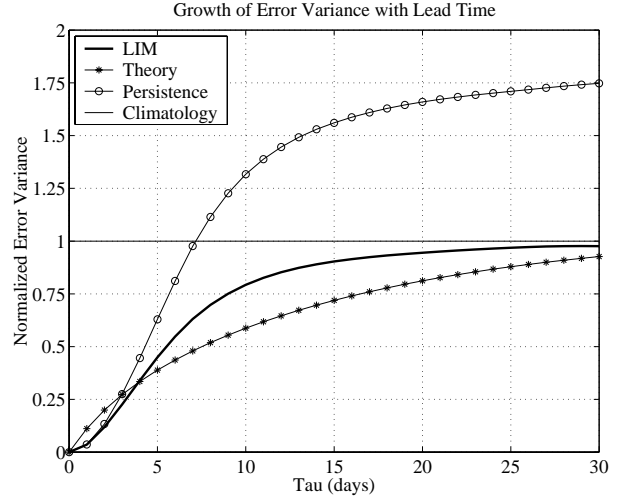

Figure 8: Thick solid line: Normalized error variance $\delta(\tau)$ of **x** forecasts based on (9) as a function of forecast lead time. Asterisks: Normalized error variance predicted by (10). Open circles: Normalized error variance of persistence forecasts. Thin solid line: Normalized error variance of climatology forecasts.

$\langle \mathbf{x}(0)\sigma(\tau)^T\rangle \neq 0$ (see also DelSole 2000). The discrepancy is also consistent with the presence of weak nonlinearity. Strong nonlinearity would, however, produce very different error growth curves (Penland and Sardeshmukh 1995, Fig. B1). Seasonal cycle effects and/or the exclusion of an important independent variable might also be factors.

b. Sensitivity to lag and EOF truncation

Although the LIM reasonably passes the Tau-test, it is clear that there is some weak dependence upon τ_0. Complicating matters further, appropriate values of τ_0 are strongly dependent upon the EOF truncation of **x** (DelSole and Hou 1999). Thus, we would like to objectively select suitable values of τ_0 and truncation, while also assessing the sensitivity of our model to these choices.

Figure 9 displays the mean pattern correlation in the North Pacific-North American region 0--90N, 120E--60W of the predicted and observed 250 hPa streamfunction anomalies at
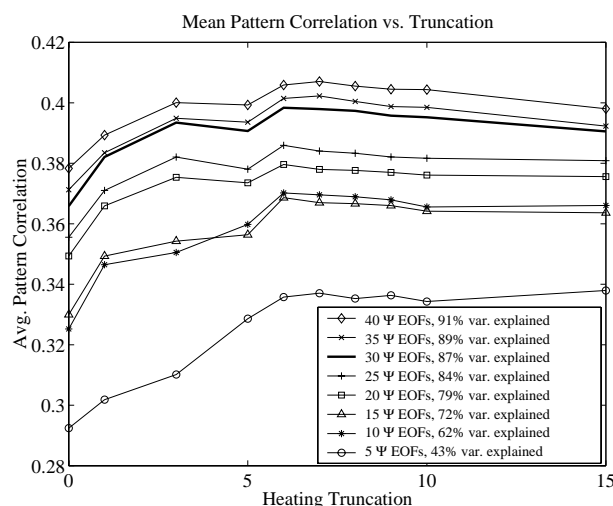
Mean Pattern Correlation vs. Truncation

Figure 9: Mean week 2 pattern correlation of 250 hPa streamfunction forecasts in the region 0--90N, 120E--60W, as a function of EOF truncation.

τ=14 d, as a function of EOF truncation. Other measures of forecast skill show similar behavior. Forecast skill is fairly flat for increasing τ_0 (not shown), although aliasing begins to arise for τ_0= 8-10 d, depending upon EOF truncation. **B** computed using 40 ψ EOFs gives the best forecasts. However, this truncation compromises analysis of the model's dynamics, in that it yields a **B** with several aliased eigenvalues and a **Q** that is not positive definite (Penland and Sardeshmukh 1995; DelSole and Hou 1999). **B** computed using 30 ψ and 7 **H** EOFs, for τ_0 = 5 d, still has high forecast skill but also has no aliasing problems. Also, only two of the 30 **Q** matrices obtained from jack-knifed estimates of **B** with these parameters are not positive definite. In both cases this is a minor problem, since only one eigenvalue for each matrix is negative and is two orders of magnitude smaller than the smallest positive eigenvalue.

It is worth stressing that the results presented in this paper are robust, in that they are qualitatively similar when any combination of 15—40 streamfunction EOFs and 5—15 heating EOFs is retained in **x**. Also, including **H** as an independent variable increases model skill,

particularly at long forecast leads (not shown). For example, a model truncation of 20 ψ and 6 **H** EOFs is about as skillful as a model constructed from 40 ψ EOFs alone (Fig. 9; see also Lo and Hendon 2000). Retaining 7 heating EOFs, however, is necessary to capture both ENSO and MJO variability. Perhaps not coincidentally, retaining 7 heating EOFs also provides the best streamfunction forecasts for most streamfunction truncations.

c. Monte Carlo tests of forecast skill

Cross validation does not eliminate the possibility that LIM skill is due to chance. As a test, Monte Carlo experiments were performed in which forecast initial conditions were randomized. These forecasts should have zero skill since the predictors and predictands are not related. We thus randomly selected (with replacement) 2280 initial conditions, made 14 day forecasts, and verified them against observations (in the original order) from Dec. 15-Feb. 28 each year. This procedure was repeated 1000 times.

The Monte Carlo results are compared to the Week 2 (cross-validated) LIM skill for each PC in Fig. 10. Mean skill of all 1000 Monte Carlo realizations (not shown) is indeed zero. We determine a 99% confidence level as the highest forecast skill reached in 1% of the realizations. This is shown separately for each PC, and never is larger than 0.055. Clearly LIM skill is not a statistical artifact.

It is still possible, however, that the LIM skill is due to persistent seasonal anomalies, especially during ENSO winters. We therefore performed a second Monte Carlo test, but this time each initial condition was randomly selected from only the observed initial conditions in that year. This also meant that the distribution of El Niño/La Niña/normal years in each 30-year sample was the same as
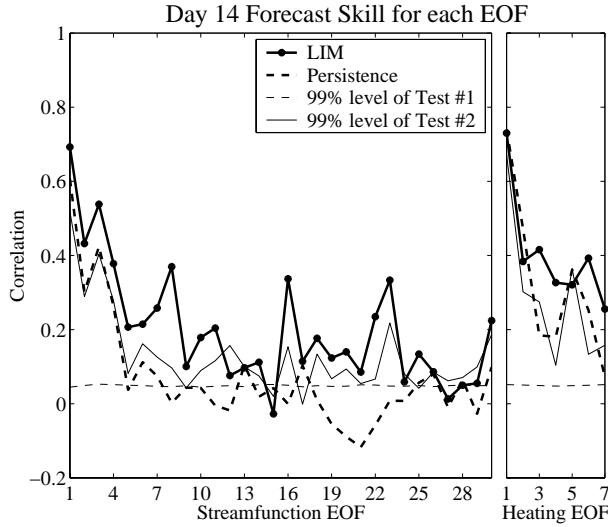
Day 14 Forecast Skill for each EOF

Figure 10: Anomaly correlation of week 2 forecasts by EOF. Thick solid line: LIM. Thick dashed line: Persistence. Thin dashed line: 99% confidence level of 1000 Monte Carlo tests (#1). Thin solid line: 99% confidence level of 1000 Monte Carlo tests, for which data is randomized only within each year (#2). See text for details.

observed. The 99% skill level for 1000 realizations of this test is also shown in Fig. 10, as is the mean skill of persistence forecasts, and again the LIM skill is shown to be real.

d. Summary

Tests for the validity of linear dynamics are met with reasonable success and in general the LIM results are robust. Furthermore, section 4 showed that LIM forecast skill is good. The success of these tests implies that diagnosis of the linear dynamics of **B** would be meaningful.

6. Dominant growth structures and the importance of tropical heating

The impact of tropical heating upon streamfunction variability is illustrated by switching off the heating feedback (that is, the heating term in the streamfunction tendency equation). We define a new linear operator $\hat{\mathbf{B}}$ which is identical to **B** except that $\mathbf{B}_{\psi\mathbf{H}}=0$ in (8). The skill of jack-knifed forecasts using $\hat{\mathbf{B}}$ is shown

in Fig. 1e. Comparing with Fig. 1f, we see that removing the heating feedback upon streamfunction reduces forecast skill everywhere. This impact is greatest in the east Pacific, reaching differences as high as 0.45 in a large region extending from southeast of Hawaii to the Gulf of Alaska. Likewise, forecast amplitude (not shown) is halved in the east Pacific when heating effects are excluded. On the other hand, comparing Fig. 1e with Fig. 1c suggests that the LIM's implicit parameterization of nonlinearity (e.g., the difference between $\mathbf{B}_{\psi\psi}$ and the linearized baroclinic operator) also leads to generally improved forecasts, particularly over the west Pacific and subtropical Atlantic.

As one might expect, the impact of tropical heating upon extratropical forecast skill increases with lead time. This is demonstrated in Fig. 11, which shows average pattern correlation as a function of forecast lead time. Results are shown for the full LIM, LIM without heating feedback upon streamfunction ($\mathbf{B}_{\psi\mathbf{H}}=0$), LIM without streamfunction feedback upon heating ($\mathbf{B}_{\mathbf{H}\psi}=0$), LIM constructed only from streamfunction, and the autocorrelation function derived from the truncated data. Note that skill is not 1.0 at day 0 because the forecasts are made in the truncated EOF space and are verified against the untruncated observations.

By Day 21, the full model's skill is almost twice that of forecasts made without diabatic forcing (cf. solid and starred curves in Fig. 11a). One may alternatively define a period of skill as the forecast lead time at which skill falls below a specified threshold. For a threshold of 0.4, the full LIM is skillful up to 14 d, or 3.5 d longer than with the heating feedback removed. The LIM constructed from streamfunction alone has more skill than the LIM in which the heating feedback is removed. This is because the ψ-only LIM parameterizes that

(a) Mean Skill of 250 mb Streamfunction Forecasts
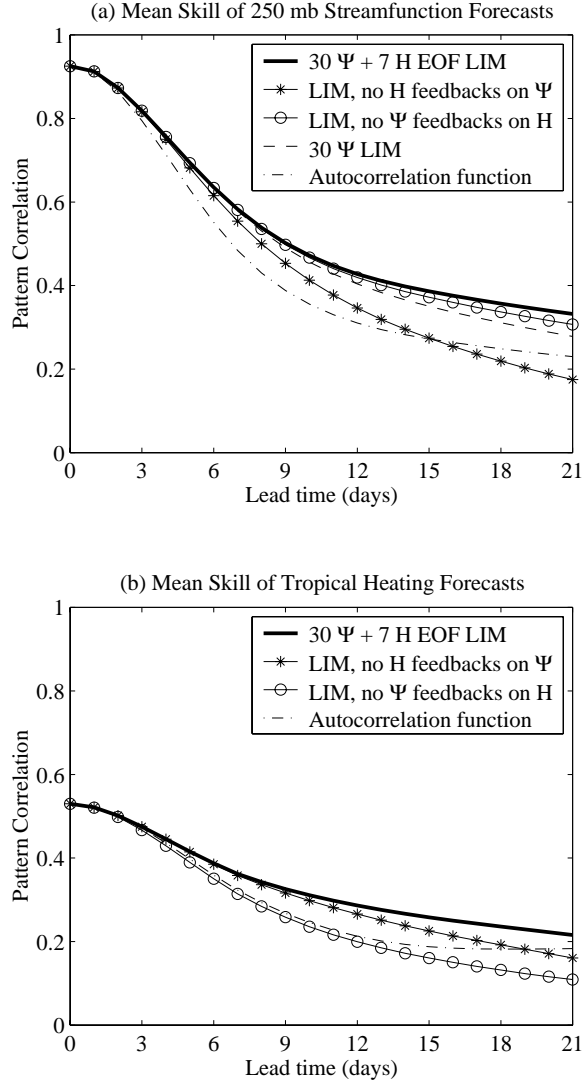

(b) Mean Skill of Tropical Heating Forecasts

Figure 11: Forecast skill as a function of lead time for the LIM and some permutations of it. The autocorrelation function (in EOF space) is also included. See text for complete details.

part of **H** linearly related to ψ. Streamfunction feedback upon heating has a notable impact upon the heating forecasts (Fig. 11b), but these degraded heating forecasts do not immediately impact streamfunction evolution.

To better understand how heating affects the streamfunction variability, we next examine structures leading to rapid streamfunction anomaly growth. Since **B** is stable, no eigenmode instability is possible. In fact, even the

least damped mode decays about 20% over 14 days. However, finite-time anomaly growth *is* possible via modal interference. Over a time interval $t=[0, \tau]$ the expected linear evolution of any initial anomaly $\mathbf{x}(0)$ is

$$\mathbf{x}(\tau)=\mathbf{G}(\tau)\mathbf{x}(0). \tag{11}$$

[To simplify notation, starting with eq. (11) we express $[\mathbf{G}(\tau_0)]^{\tau/\tau_0}$ as $\mathbf{G}(\tau)$.] Using the SVD of $\mathbf{G}(\tau) = \mathbf{U}\mathbf{S}\mathbf{V}^T$, this becomes

$$\mathbf{U}^T\mathbf{x}(\tau)=\mathbf{S}\mathbf{V}^T\mathbf{x}(0). \tag{12}$$

U and **V** form complete orthonormal spaces spanned by the set of singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, respectively, with associated singular values s_i. Singular vector pairs describe the evolution of $\mathbf{x}$ over the optimization interval τ, in the sense that $\mathbf{v}_i$ evolves into $s_i\mathbf{u}_i$. If $\mathbf{x}$ is projected into the space of **V** as

$$\mathbf{x}(0) = \sum_i \beta_i \mathbf{v}_i,$$

where β_i are the projection coefficients, then from (12),

$$\mathbf{x}(\tau) = \sum_i s_i \beta_i \mathbf{u}_i.$$

The magnitude of anomaly growth over this time interval is

$$\frac{E(\tau)}{E(0)} = \frac{\langle \mathbf{x}(\tau), \mathbf{x}(\tau) \rangle}{\langle \mathbf{x}(0), \mathbf{x}(0) \rangle} \tag{13}$$

$$= \frac{\langle \mathbf{G}(\tau)\mathbf{x}(0), \mathbf{G}(\tau)\mathbf{x}(0) \rangle}{\langle \mathbf{x}(0), \mathbf{x}(0) \rangle}$$

(e.g., Farrell 1988, 1989). Here E represents any inner vector product $<\bullet,\bullet>$ in the norm **D**; the SVD of $\mathbf{G}(\tau)$ is performed under this norm. Thus, for an initial structure equivalent to the right singular vector $\mathbf{v}_i$, the globally integrated measure E will change by a factor of $\lambda_i=s_i^2$ after τ days. Anomaly *growth* occurs when

$x(0)$ projects strongly on right singular vectors whose associated singular values s_i are greater than unity. The maximum amplitude of this growth, λ_1, occurs if $\mathbf{x}(0)$ is proportional to the leading, or 'optimal,' right singular vector $\mathbf{v}_1$.

Because we are primarily interested here in the growth of ψ, and not of $\mathbf{H}$, we let $\mathbf{D}$ represent an L2 norm of streamfunction only; in PC space $D_{ij}=1$ if $i = j \leq 30$, and otherwise $D_{ij}=0$. A plot of λ_1 as a function of τ (Fig. 12), referred to as the maximum amplification (MA) curve (Penland and Sardeshmukh 1995), peaks at 19 days with a growth factor of 3.5. Optimal growth is however possible out to 70 days if the initial state is identical to $\mathbf{v}_1$. The importance of heating can be seen by comparing the MA curve of $\mathbf{G}$ with that computed using

$$\hat{\mathbf{G}} = \exp(\hat{\mathbf{B}}\tau)$$

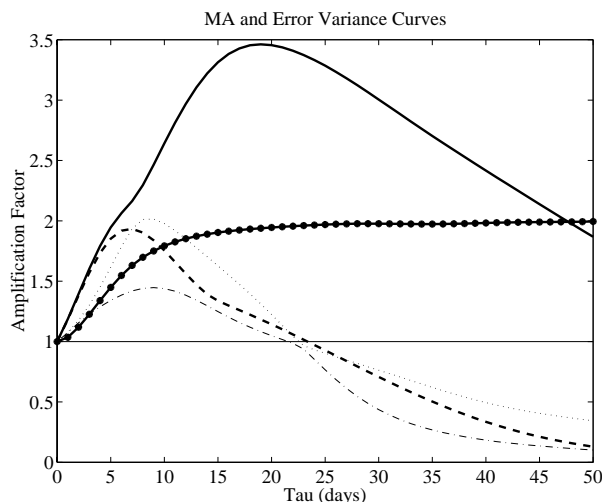(dashed curve in Fig. 12). For $\hat{\mathbf{G}}$, streamfunction growth maximizes at only 7 days and is

Figure 12: Comparison of MA curves with growth of error variance. Solid line: MA curve (λ_1). Dashed line: MA curve for $\mathbf{B}_{\psi\mathbf{H}}=0$; that is, when the effect of heating on streamfunction is eliminated from $\mathbf{B}$. Dotted line: Amplification curve for second singular vector (λ_2). Dash-dot line: Amplification curve for third singular vector (λ_3). Filled circles: Growth of actual error variance $1 + \delta(\tau)$.

not possible beyond day 23. A similar lack of long-term growth is evident in unforced barotropic model simulations (Sardeshmukh et al. 1997). The MA curves thus show the crucial role of tropical diabatic heating in extratropical anomaly growth at the medium and extended ranges. For comparison the curves representing amplification factors λ_2 and λ_3 associated with $\mathbf{v}_2$ and $\mathbf{v}_3$ are also presented in Fig. 12.

Streamfunction anomalies can thus undergo long-term growth even in this stable system. However, for this growth to be predictable the anomalies must amplify above the noise (see (3)). A predictability limit may be defined by comparing the predictable and unpredictable anomaly growth for an initial perturbation of unit amplitude. The MA curve shows maximum predictable growth (that is, due to the effects of $\mathbf{B}$), and can be read as the amplification factor of an optimal initial perturbation with amplitude normalized to 1. Unpredictable anomaly growth (that is, error growth) is given by $1 + \delta(\tau)$ (cf. Fig. 8 and section 5a), plotted as the line with filled circles in Fig. 12. The point at which these two curves intersect is the lead time at which the error growth (on average) becomes equal to the maximum possible predictable anomaly growth, and can therefore be interpreted as a useful predictability limit (see also Penland and Sardeshmukh 1995). Fig. 12 shows that weekly averages are potentially predictable in this sense for almost fifty days. When tropical diabatic forcing is excluded, however, predictability is limited to about ten days. Even the second singular vector of $\mathbf{B}$ (dotted line) has greater predictability than this. These results are consistent with numerical model experiments indicating a degradation of forecast skill when tropical forcing is inadequately represented (e.g., Sardeshmukh and Hoskins 1988; Ferranti et al. 1990; Klinker 1990; Cai et al. 1996; Hendon et al. 2000).

a) Day 0 Ψ_{250}

b) Day 14 Ψ_{250}

c) Day 0 **H**

d) Day 14 **H**

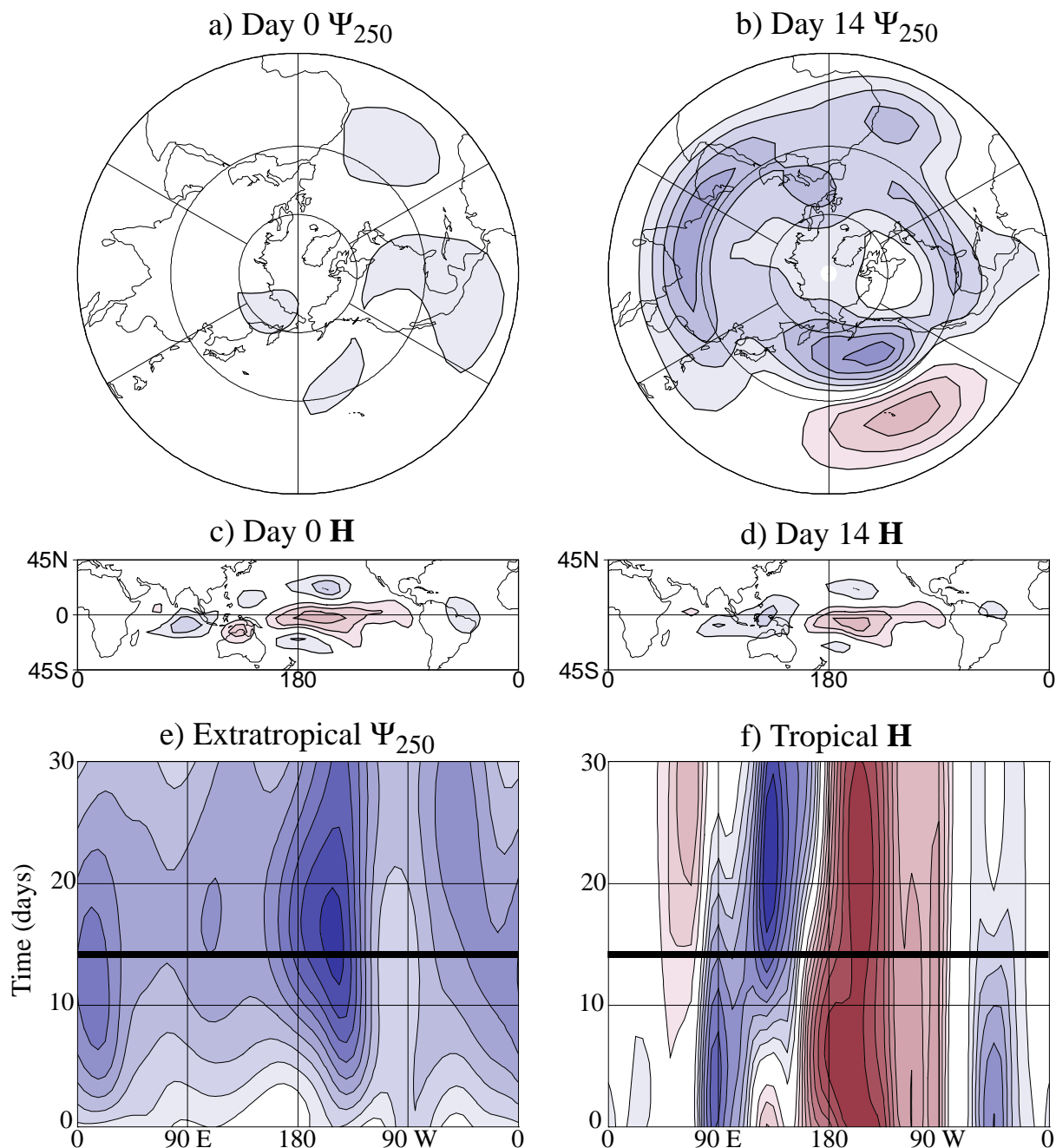### e) Extratropical Ψ_{250}

f) Tropical **H**

Figure 13: Leading singular vector pair. (a) Initial condition of 250 hPa streamfunction. (b) Optimal growth structure of 250 hPa streamfunction into which the initial condition grows 14 days later. (c) Initial condition of tropical heating. (d) Tropical heating 14 days later. A heating contour interval of 100 W m^{-2} corresponds to streamfunction contours of 9.8 x 10^6 m^2 s^{-1}. Positive values are denoted by red shading; negative values are denoted by blue shading. At τ =14 days λ_1 = 3.22 . (e) Hovmuller diagram of 30°N--50°N averaged 250 hPa streamfunction. Contour interval is half that in (a) and (b). (f) Hovmuller diagram of 20°S--5°N averaged tropical diabatic heating. Contour interval is half that in (c) and (d). The horizontal thick black line indicates Day 14; Figs. b and d correspond to this time.

For an optimization interval of τ=14 d, the three leading singular vectors all grow. The optimal initial and evolved structures associ-ated with maximum growth, $\mathbf{v}_1$ and $\mathbf{G}(14)\mathbf{v}_1$, are shown in Figs. 13a-d. (Results at 750 mb indicate roughly equivalent barotropic struc-

tures and are not shown.) Hovmuller diagrams depicting the evolution of extratropical 250 mb streamfunction and tropical heating from the initial condition $\mathbf{v}_1$ are shown in Figs. 13e and f. Although our choice of norm maximizes only streamfunction growth, anomalous heating can and does evolve. The optimal initial condition for growth has little coherent streamfunction information but has strong anomalous heating in the central Pacific and along the northwest coast of Australia. By day 14 the streamfunction anomaly has strongly amplified and evolved into a pattern with a strong zonally symmetric component, plus a deep equivalent barotropic low off the Aleutians with evidence of wave propagation from the Tropics (Fig. 13b). This evolution occurs whether or not the initial weak streamfunction anomaly is present. In fact, this structure depends upon anomalous heating throughout its evolution. If the heating is turned off at any point, global streamfunction growth ceases. Similarly, we find that an anomaly consisting of the east Pacific portion of Fig. 13b (the north-south dipole) immediately decays in the absence of heating. Moreover, there is no equivalent structure resulting from the SVD of $\hat{\mathbf{G}}$ (14). These results all suggest that the streamfunction pattern of Fig. 13b, which is very similar to the leading EOF, undergoes predictable growth only in the presence of tropical forcing. Even so, this streamfunction response is not purely a tropically-forced linear baroclinic Rossby wave, since its evolution is still affected by the parameterized nonlinearities within the full $\mathbf{B}$.

The evolution of tropical heating (Fig. 13f) consists of an MJO propagating through the tropical IndoPacific in the presence of a relatively steady central Pacific ENSO heating anomaly. Such behavior in the leading singular vector is seen for all optimization intervals greater than nine days. In all these cases, the strongest response (and in particular, the maximum zonal jet anomaly in the eastern Pacific)

occurs as the MJO heating anomaly propagates to be in phase with the ENSO heating anomaly (Fig. 13f). As τ is increased, the MJO component of $\mathbf{v}_1$ is weakened relative to its ENSO component (not shown), suggesting that a stronger MJO during an ENSO event could produce a more rapidly amplifying but less persistent streamfunction response. Additional calculations in which the tropical heating is not allowed to evolve (that is, $\mathbf{H}$ is kept fixed at its initial value) show that the heating evolution is particularly important to streamfunction evolution after two weeks. In fact, the evolving heating contributes significantly to the ultimate *weakening* of the streamfunction anomaly. Over shorter time intervals, however, most of the streamfunction evolution can be reproduced by holding the initial heating anomaly fixed. Similar behavior is seen for the other two singular vectors discussed below. Given the slower time scales associated with the tropical heating (at least, as represented by the leading 7 EOFs), this result is not surprising. It also suggests that poor tropical heating forecasts (as opposed to a poor initialization of the heating) will have a steadily increasing impact on forecast skill. For example, LIM skill is not obviously higher than the DERF skill until Week 3 (see Fig. 4), which would be expected if it took a few weeks for the poor DERF tropical heating forecasts to influence the extratropical anomalous streamfunction.

The results of Figs. 12 and 13 are robust. Qualitative features of these plots are insensitive to the heating dataset, the temporal filter, the EOF truncation of $\mathbf{x}$, and the choice of τ_0. Note also that this optimal structure projects strongly upon many different eigenmodes and not just the least damped mode. Thus this growth cannot be identified with that of any one (or even a few) eigenmode(s).

The growth indicated in Fig. 13 is, in principle, only possible. Does such growth actually

occur? And perhaps more important, is the observed growth linearly proportional to the projection of the initial anomaly on the leading right singular vectors; that is, does the system act in a truly linear manner? To answer this, observed low-frequency anomalies were projected on the initial condition (e.g., Figs. 13 a, c), and compared to the projection of the observed streamfunction upon the predicted evolved structure (e.g., Figs. 13b, d) 14 days later. A scatterplot of the results is shown in Fig. 14a. Positive slope between these projections, high linear correlation (0.78), and small scatter about the least-squares line indicates that optimal growth often occurs in the way shown in Fig. 13 in the real atmosphere. Of course, some scatter about the least-squares line is expected due to the noise $\mathbf{F}_s$. Nevertheless, the good fit and high correlation are consistent with (2) being a good model of extratropical low-frequency variability. Similar calculations for the other two growing structures are displayed in Figs. 14b and c. Note that the slope is less for the second singular vector and less still for the third. The slopes of these lines should correspond to the amplification factors λ_i, and in fact they are quite close. This also serves as additional evidence that $\mathbf{B}$ is independent of τ_0, since these calculations have all been made at τ=14 d even though $\mathbf{B}$ was computed using τ_0=5 d.

The second and third growing singular vectors for τ=14 d are displayed in Figs. 15 and 16, respectively. The second vector (Fig. 15) evolves into a strong wavetrain dispersing into the extratropics in response to both MJO-like tropical heating and an initial perturbation in the southeast Asian jet. Unlike the leading singular vector, the initial streamfunction anomaly is also important for growth. Initializing with only the streamfunction component, or only the heating component, of $\mathbf{v}_2$ results in similar wavetrains (not shown). Also, this pattern is similar to the second singular vector

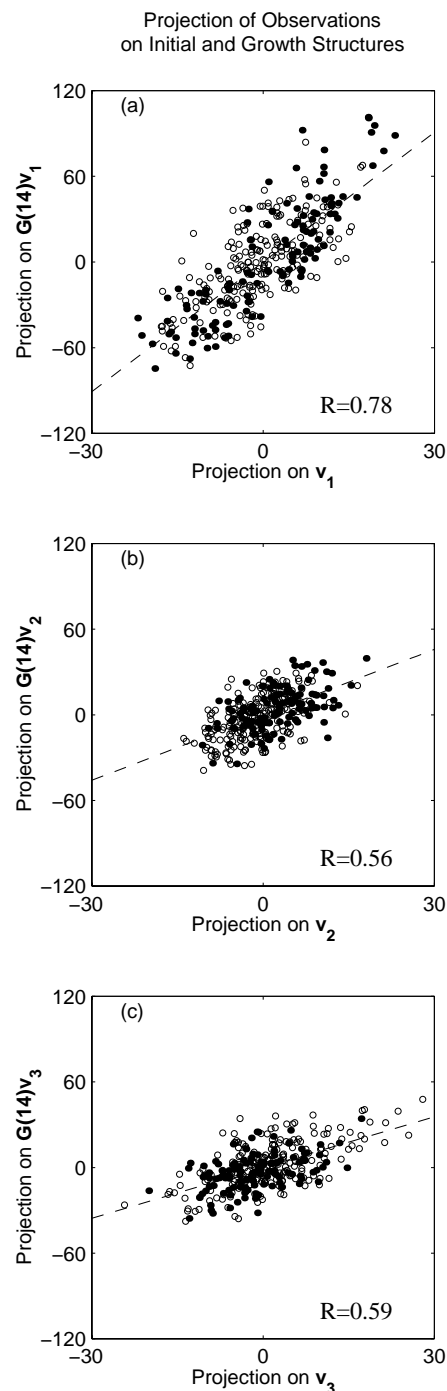Projection of Observations
on Initial and Growth Structures

Figure 14: Scatterplot of the projection of observations on $\mathbf{v}_i$ versus the projection of observations on $\mathbf{G}(14)\mathbf{v}_i$ fourteen days later, for i=1, 2, or 3. Dashed line: Least-squares fit to the data. Open circles: Non-ENSO winters. Filled circles: ENSO winters. For clarity, values are only plotted once every seven days. The strong positive slope and high correlation indicates that optimal growth does occur in the atmosphere.
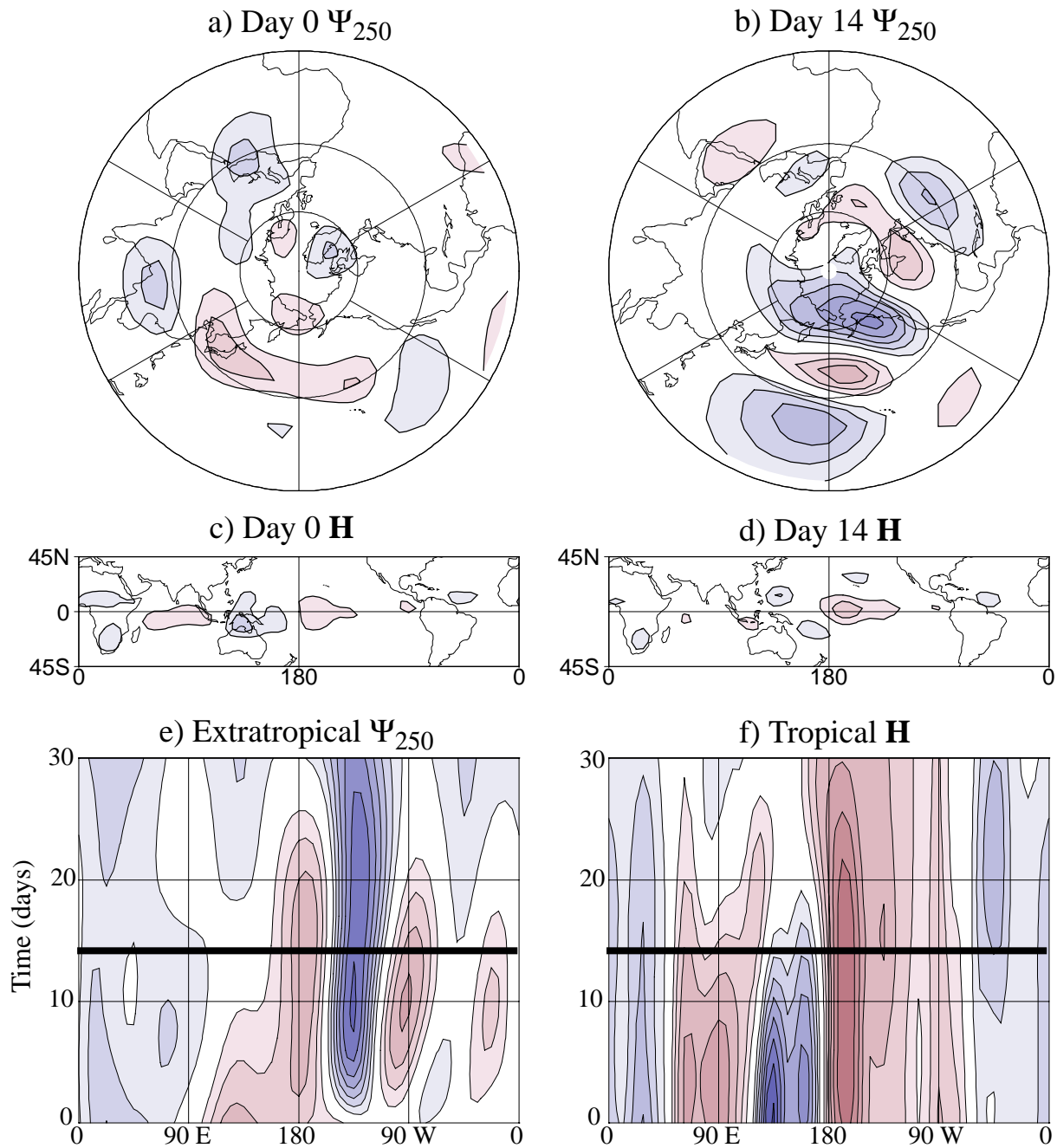
a) Day 0 Ψ_{250}

b) Day 14 Ψ_{250}

c) Day 0 **H**

d) Day 14 **H**

e) Extratropical Ψ_{250}

f) Tropical **H**

Figure 15: As in Fig. 13, but for the second singular vector pair. At 14 days $\lambda_2 = 1.70$.

pair of $\hat{\mathbf{G}}(14)$, although the latter exhibits much weaker growth over the 14-day period. This suggests that the role of the initial heating perturbation in the dynamics of the second singular vector is mainly to amplify a structure that can grow even in its absence.

No matter what the optimization interval is for this second singular vector, maximum stream-function amplitude occurs after the Indian Ocean heating anomaly has weakened and then shifted into Indonesia (Fig. 15f). A similar mechanism was noted in the composite study of Higgins and Mo (1997). However,
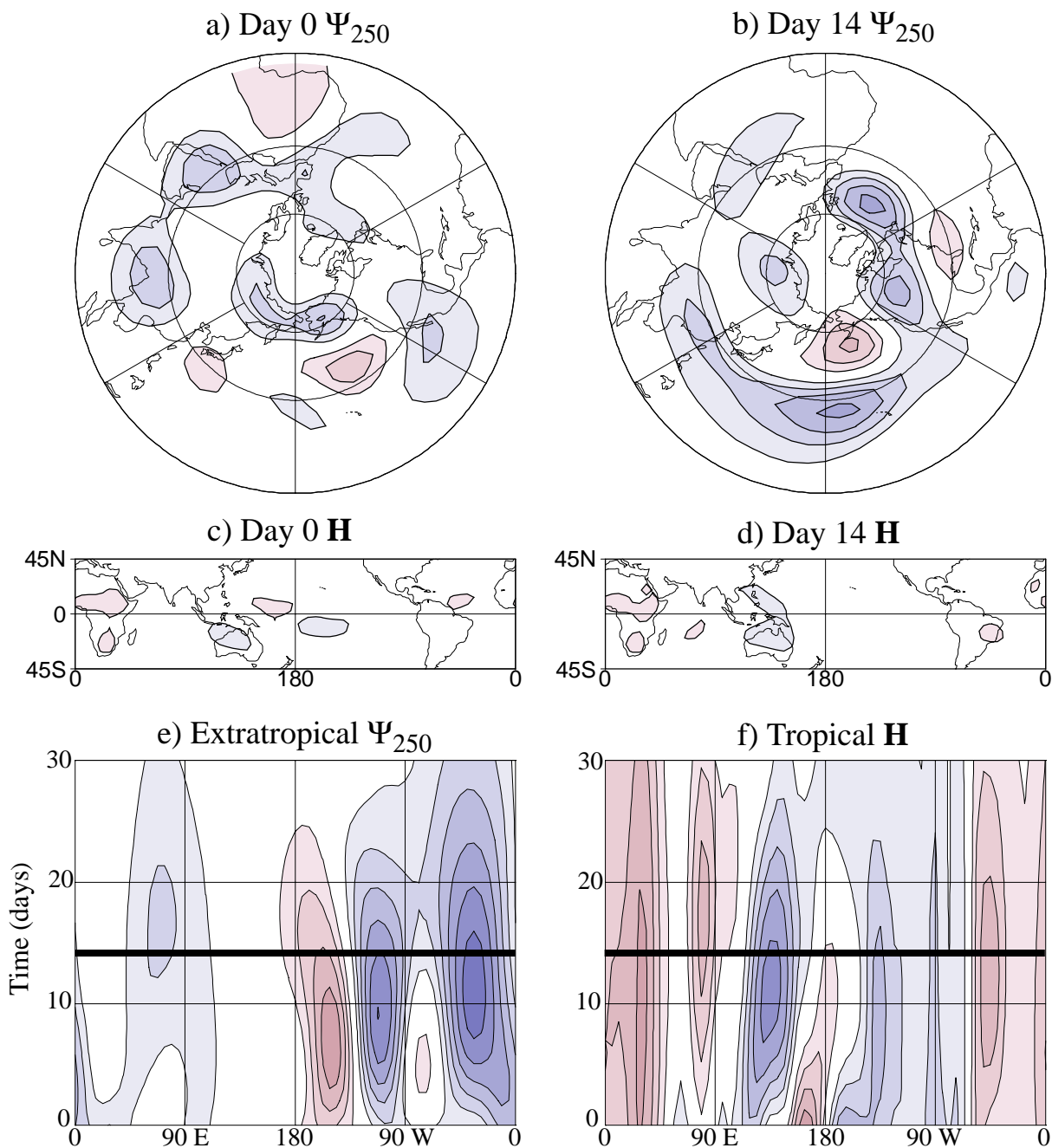
a) Day 0 Ψ_{250}

b) Day 14 Ψ_{250}

c) Day 0 **H**

d) Day 14 **H**

e) Extratropical Ψ_{250}

f) Tropical **H**

Figure 16: As in Fig. 13, but for the third singular vector pair. At 14 days $\lambda_3 = 1.30$.

although global growth is maximized for an optimization interval of $\tau = 9$ days, the *regional* Pacific basin anomaly is considerably larger for $\tau = 21$ days (not shown). As τ is increased, the initial tropical heating anomaly becomes larger relative to the initial streamfunction perturbation. Consequently, the

streamfunction anomaly becomes both stronger and more persistent in the Pacific south of 45°N, but has relatively reduced amplitude over Alaska and much less downstream development. This suggests that intense North Pacific low-frequency anomalies of the type associated with anomalous precipitation in the

Western United States (e.g., Mo and Higgins 1998) can be initiated without tropical forcing, but will be significantly more persistent with tropical forcing.

The third singular vector (Fig. 16) evolves into a PNA-like pattern which is very similar to the second EOF. The evolution of this structure shows more pronounced downstream energy dispersion than the first two singular vectors (see Fig. 16e). Tropical heating is still present but is of secondary importance for the development of this anomaly. When a 14-day forecast is initialized with only the streamfunction portion of $\mathbf{v}_3$, an anomaly similar to the evolved streamfunction structure (Fig. 16b) is produced, but with about 70% of its amplitude. The leading singular vector of $\hat{\mathbf{G}}(14)$ is also very similar to the streamfunction component in Fig. 16. These results are consistent with earlier analyses of GCM output (Lau 1983; Branstator 1992; Feldstein 1998), which suggested that PNA-like anomalies are primarily due to internal extratropical dynamics, rather than being tropically forced. Still, it is worth noting that for optimization intervals greater than 14 days, tropical forcing becomes increasingly important in the evolution of even this singular vector, particularly in the region of energy dispersion over the Atlantic.

We should stress that the forecast skill of our model is not solely due to optimal growth of the leading singular vector. Projection of the observations upon $\mathbf{v}_1$ (see Fig. 13) versus the skill of 14 day forecasts initialized with those anomalies is shown in Fig. 17. There is higher skill as the amplitude of this projection increases, but clearly good skill can exist even if the initial condition projects weakly upon the leading singular vector. On the other hand, 4% of the forecast initializations have small projection (that is, in the middle tercile) upon all three growing singular vectors, and an additional 6% have large projection only upon the
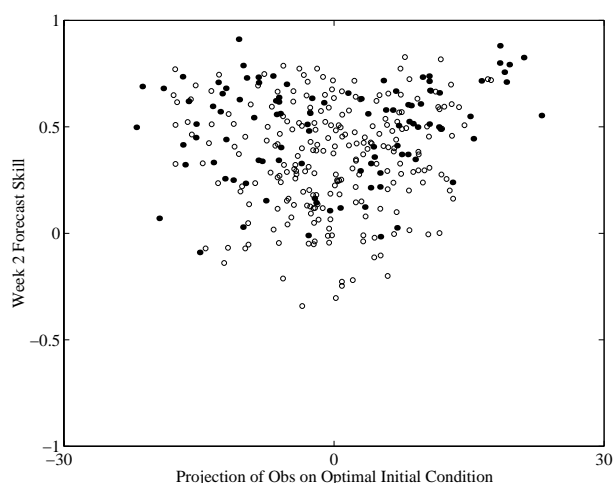

Figure 17: Scatterplot of the spatial pattern correlation of week 2 250 hPa streamfunction forecasts versus the projection of observations on the leading right singular vector (see Fig. 13a and c). Open circles: non-ENSO winters. Filled circles: ENSO winters. For clarity, values are only plotted once every seven days.

third singular vector. Forecasts at these times generally result in the LIM's poorest skill (with average pattern correlation of about 0.22 at day 14).

Note that the initial heating perturbation in the leading vector is dominated by heating in the central Pacific. Does this mean that optimal growth is mainly due to ENSO forcing, and most of the predictability is confined to periods of large tropical SST anomalies? To answer these questions, the results shown in Figs. 14 and 17 are segregated into ENSO (filled circles) and non-ENSO years (open circles). ENSO years are defined as those winters in which the magnitude of the December, January, and February Niño 3.4 SST anomaly exceeds one standard deviation (as determined from 1958—1999 wintertime monthly mean SST). Four El Niño and five La Niña winters can be so defined for the 30 year record used in this study (Table 1). [Defining ENSO by month rather than by season, or by using an index based on combining Niño 3.4 with SOI (Smith and Sardeshmukh 2000), gives nearly identical results.] Figure 14 shows that heating

Table 1: ENSO winters considered in this paper. El Niño (La Niña) winters are defined as winters for which December, January, and February SST is greater (lesser) than the 1958—1999 monthly mean SST by one standard deviation.

El Niño winters	La Niña winters
1972—1973	1970—1971
1982—1983	1973—1974
1991—1992	1975—1976
1997—1998	1988—1989
	1998—1999

patterns similar to Fig. 13a occur in both ENSO and non-ENSO years. Similarly, Fig. 17 shows that forecasts during ENSO years are only a little more skillful on average than forecasts during non-ENSO years, primarily due to fewer poor forecasts. These results suggest tropical diabatic heating, not tropical sea surface temperature *per se*, is more directly relevant to medium and extended range extratropical predictability.

7. Concluding remarks

A linear model of Northern Hemisphere wintertime low-frequency variability has been constructed from the observed statistics of weekly-averaged streamfunction and tropical diabatic heating anomalies. The model produces medium-range forecasts with comparable skill to MRF ensemble-mean predictions. The fact that the model is significantly more skillful than models based on some simple linearized dynamical theories of low-frequency variability points to serious deficiencies in those theories. The fact that the model skill is comparable to that of the NCEP operational model suggests, at the very least, that the LIM should provide a useful predictability benchmark for GCMs in the extended medium

range. This skill also justifies our use of the LIM to diagnose and understand low-frequency variability.

Two primary reasons exist for the good LIM forecast skill. First, the empirical-dynamical operator $\mathbf{B}$ implicitly includes a linear representation of nonlinear processes absent in the linearized governing equations, such as the feedback of synoptic eddies on low-frequency flow. Second, the LIM allows tropical heating to evolve and force the extratropical circulation. Neglecting diabatic forcing degrades forecast skill nearly everywhere, especially in the southeastern United States, tropical Atlantic, and eastern Pacific. It is important to note that we determined this impact not by relating the change in forecast skill to increased heating degrees of freedom in the model (as one might do, for example, by interpreting Fig. 9), but rather by using the LIM framework to remove the effects of heating (through $\mathbf{B}_{\psi/\mathbf{H}}$). This explicit treatment of heating allowed evaluation of the heating impact upon forecast skill.

Diagnosis of the LIM indicates that the structure and magnitude of tropical diabatic heating is crucially important in the long-term (>10 day) evolution of low-frequency circulation anomalies. Indeed, the principal predictable growth structure of the atmosphere (which is similar to the leading EOF of streamfunction) must be diabatically forced throughout its evolution. Other structures, however, may be more dependent on parameterized nonlinearities such as transient eddy feedback and other unresolved baroclinic processes. Also, there may be some atmospheric states which are more predictable than others (see Fig. 12). This suggests that tropical forcing may affect predictability in a state-dependent manner, a point further explored in Newman et al. (2001a, submitted to *Quart. J. Roy. Meteor. Soc.*).

We conclude that the dynamics of low-frequency variability are effectively linear, stable, and stochastically forced. Because some tests of linearity meet with only fair success, weak unparameterizable nonlinearity may exist, especially over the North Atlantic, but not enough to invalidate the fundamental linear hypothesis. Also, further improvement of the linear model is certainly possible. Most notably, our $\mathbf{x}$ may not represent all the variables (and/or sufficient vertical resolution) necessary for a linear parameterization of the nonlinear feedbacks. For example, some details of the vertical structure of the heating (especially in the subtropics) and some tropopause and/or surface variable(s) might provide additional information that cannot be represented as a linear combination of the current model variables. Including seasonal cycle effects in $\mathbf{B}$ and accounting for the finite decorrelation time scale of the noise might also improve the model and its forecasts.

Given the evidence that the LIM is a good forecast model, we plan to use it to develop a general model of extratropical low frequency variability. In particular, we will assess the extent to which *predictable* low-frequency variability in the extratropics is predominantly the result of low-frequency tropical diabatic heating which forces extratropical anomalies that are additionally modified by transient eddy feedbacks. The zonally asymmetric time-mean basic state is important in this view, but not only or even mainly because it acts as an energy source/sink for the eddies. Instead, it acts to organize variability in three ways: (1) by localizing the Rossby wave source due to heating in regions of strong upper-tropospheric potential vorticity gradients (Sardeshmukh and Hoskins 1988); (2) by localizing the storm track, and thus the transient eddy feedback (Farrell and Ioannou 1993; Whitaker and Sardeshmukh 1998); and (3) by steering the low-frequency anomalies along Rossby

waveguides (Hoskins and Karoly 1981; Simmons 1982; Hoskins and Ambrizzi 1993; Borges and Sardeshmukh 1995).

Our LIM analysis provides a framework in which the relative importance of the above mechanisms can be *quantified*. This is the main goal of Part II (Newman et al. 2001b). Part of the analysis has already been carried out here by demonstrating the impact of tropical heating, through the $\mathbf{B_{\psi H}}$ term, on both forecast skill and anomaly growth. In Part II, we will compare the submatrices in $\mathbf{B}$ (especially $\mathbf{B_{\psi\psi}}$) with the linear operators associated with different dynamical models. In particular, if $\mathbf{B_{\psi\psi}} = \mathbf{L} + \mathbf{T_{\psi\psi}}$ then we can determine $\mathbf{T_{\psi\psi}}$ given $\mathbf{L}$ (associated with, say, a linear baroclinic model). Note that it is considerably more difficult to perform such an analysis from just MLR or related techniques such as CCA. In those cases, the effects of $\mathbf{L}$ and $\mathbf{T_{\psi\psi}}$ are mixed together with the effects of tropical heating in $\mathbf{G}(\tau_0)$. MLR alone also provides an incomplete analysis of the noise term, since although the forecast error can be determined from MLR, this error is the result of a convolution of the noise with the propagator operator (see eq. (3b)).

Finally, we reiterate that although our LIM yields the correct dynamical operator of low-frequency evolution, it does not amount to a complete dynamical theory, nor does it replace the need for such a theory. That is, we have not provided a physical basis for why the elements of $\mathbf{B}$ are what they are. Nevertheless, our LIM should provide important constraints on any comprehensive dynamical theory of low-frequency variability. Such a theory should be consistent with the form of (2), yield forecast models that are at least as skillful as the LIM, produce structures similar to Figs. 13, 15, and 16, and indicate a pronounced influence of tropical heating upon extratropical low-fre-

quency variability with details similar to those shown here.

Acknowledgements

References

Borges, M. D., and P. D. Sardeshmukh, 1995: Barotropic Rossby wave dynamics of zonally varying upper level flows during northern winter. *J. Atmos. Sci.*, **52**, 3779-3796.

Branstator, G. W., 1992: The maintenance of low-frequency anomalies. *J. Atmos. Sci.*, **49**, 1924-1945.

Branstator, G. W., 1995: Organization of storm track anomalies by recurring low-frequency circulation anomalies. *J. Atmos. Sci.*, **52**, 207-226.

Branstator, G. W., and S. E. Haupt, 1998: An empirical model of barotropic atmospheric dynamics and its response to tropical forcing. *J. Climate*, **11**, 2645—2667.

Cai, M., J. S. Whitaker, R. M. Dole, and K. L. Paine, 1996: Dynamics of systematic errors in the NMC medium range forecast model. *Mon. Wea. Rev.*, **124**, 265-276.

Cherry, S., 1996: Singular value decomposition analysis and canonical correlation analysis. *J. Climate*, **9**, 2003—2009.

DelSole, T., and A. Y. Hou, 1999: Empirical stochastic models for the dominant climate statistics of a general circulation model. *J. Atmos. Sci.*, **56**, 3436—3456.

DelSole, T., 2000: A fundamental limitation of Markov models. *J. Atmos. Sci.*, **57**, 2158-2168,.

Farrell, B., 1988: Optimal excitation of neutral Rossby waves. *J. Atmos. Sci.*, **45**, 163-172.

Farrell, B., 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46**, 1193-1206.

Farrell, B., and P. J. Ioannou, 1993: Stochastic dynamics of baroclinic waves. *J. Atmos. Sci.*, **50**, 4044-4057.

Feldstein, S., 1998: The growth and decay of low-frequency anomalies in a GCM. *J. Atmos. Sci.*, **55**, 415—428.

Ferranti, L., T. N. Palmer, F. Molteni, and E. Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation and its impact on medium and extended range prediction. *J. Atmos. Sci.*, **47**, 2177-2199.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer Verlag, 442 pp.

Haltiner, G. J., and R. T. Williams, 1980: *Numerical Prediction and Dynamic Meteorology.* John Wiley & Sons, 477 pp.

Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus*, **28**, 474--485.

Hendon, H. H., B. Liebmann, M. Newman, and J. D. Glick, 2000: Medium-range forecast errors associated with active episodes of the Madden-Julian oscillation. *Mon. Wea. Rev.*, **128**, 69—86.

Higgins, R. W., and K. C. Mo, 1997: Persistent North Pacific circulation anomalies and the Tropical Intraseasonal Oscillation. *J. Climate*, **10**, 223-244.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.,* **38**, 1179-1196.

Hoskins, B. J. and T. Ambrizzi, 1993: Rossby wave propagation on a realistic longitudinally varying flow. *J. Atmos. Sci.*, **50**, 1661-1671.

Johnson, S. D., D. S. Battisti, and E. S. Sarachik, 2000: Empirically derived Markov models and prediction of Tropical Pacific sea surface temperature anomalies. *J. Climate*, **13**, 3-17.

Kiladis, G. N., and K. M. Weickmann, 1992: Circulation anomalies associated with tropical convection during Northern winter. *Mon. Wea. Rev.*, **120**, 1900-1923.

Klinker, E., 1990: Investigation of systematic errors by relaxation experiments. *Quart. J. Roy. Meteor. Soc.*, **116**, 573-594.

Lau, N.-C., 1983: Midlatitude wintertime circulation anomalies appearing in a 15-yr GCM experiment. *Large Scale Dynamical Processes in the Atmosphere*, B. J. Hoskins and R. P. Pearce, Eds., Academic Press, p. 111-125.

Lau, K.-M. and T. J. Phillips, 1986: Coherent fluctuations of extratropical geopotential height and tropical convection in intraseasonal timescales. *J. Atmos. Sci.*, **43**, 1164-1181.

Liebmann, B. and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Met. Soc.*, **77**, 1275--1277.

Lo, F., and H. H. Hendon, 2000: Empirical extended-range prediction of the Madden-Julian Oscillation. *Mon. Wea. Rev.*, **128**, 2528-2543.

Mo, K. C., and R. W. Higgins, 1998: Tropical convection and precipitation regimes in the western United States. *J. Climate*, **11**, 2404-2423.

Mo, K. C., 1999: Alternating dry and wet episodes over California and intraseasonal oscillations. *Mon. Wea. Rev.*, **127**, 2759-2776.

Newman, M., and P. D. Sardeshmukh, 1995: A caveat concerning singular value decomposition. *J. Climate*, **8**, 352—360.

Newman, M., P. D. Sardeshmukh, and C. Penland, 1997: Stochastic forcing of the wintertime extratropical flow. *J. Atmos. Sci.*, **54**, 435—455.

Newman, M., and P. D. Sardeshmukh, 1998: The impact of the annual cycle on the North Pacific/North American response to remote low-frequency forcing. *J. Atmos. Sci.*, **55**, 1336—1353.

Newman, M., M. A. Alexander, C. R. Winkler, J. D. Scott, and J. J. Barsugli, 2000: A linear diagnosis of the coupled extratropical ocean atmosphere system in the GFDL GCM. *Atmospheric Sci. Lett.*, **1**, 1-12.

Newman, M., P. D. Sardeshmukh, and C. R. Winkler, 2001a: An estimate of predictability limits on sub-seasonal timescales. Submitted to *Quart. J. Roy. Meteor. Soc.*

Newman, M., P. D. Sardeshmukh, and C. R. Winkler, 2001b: A linear model of wintertime low-frequency variability. Part II: Diagnosis of covariance statistics. To be submitted to J. Climate.

Opsteegh, J.D. and H.M. van den Dool, 1980: Seasonal differences in the stationary response of a linearized primitive equations model: Prospects for long-range weather forecasting? *J. Atmos. Sci.*, **37**, 2169-2185.

Papanicolaou, G. and W. Kohler, 1974: Asymptotic theory of mixing stochastic ordinary differential equations. *Commun. Pure Appl. Math.*, **27**, 641-668.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.*, **117**, 2165—2185.

Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D*, **98**, 534--558.

Penland, C., and M. Ghil, 1993: Forecasting Northern Hemisphere 700-mb geopotential heights using principal oscillation patterns. *Mon. Wea. Rev.*, **121**, 2355—2372.

Penland, C., and T. Magorian, 1993: Prediction of Niño-3 sea surface temperatures using linear inverse modeling. *J. Climate*, **6**, 1067—1076.

Penland, C., and L. Matrosova, 1994: A balance condition for stochastic numerical models with application to the El Niño-Southern Oscillation. *J. Climate*, **7**, 1352-1372.

Penland, C., and L. Matrosova, 1998: Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. *J. Climate*, **11**, 483—496.

Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical sea surface temperature anomalies. *J. Climate*, **8**, 1999—2024.

Qin, J. and Robinson, W. A., 1995: The impact of tropical forcing on extratropical predictability in a simple global model. *J. Atmos. Sci.*, **52**, 3895-3910.

Qin, J., and van den Dool, H. M., 1996: Simple extensions of an NWP model. *Mon. Wea. Rev.*, **124**, 277-287.

Sardeshmukh, P. D., and B. J. Hoskins, 1984: Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524—2529.

Sardeshmukh, P. D., and B. J. Hoskins, 1985: Vorticity balances in the tropics during the 1982—1983 El Niño-Southern Oscillation event. *Quart. J. Roy. Meteor. Soc.*, **111**, 261—278.

Sardeshmukh, P. D., and B. J. Hoskins, 1988: The generation of global rotational flow by steady idealized divergence. *J. Atmos. Sci.*, **45**, 1228-1251.

Sardeshmukh, P. D., 1993: The baroclinic chi problem and its application to the diagnosis of atmospheric heating rates. *J. Atmos. Sci.*, **50**, 1099—1112.

Sardeshmukh, P. D., M. Newman, and M. D. Borges, 1997: Free barotropic Rossby wave dynamics of the wintertime low-frequency flow. *J. Atmos. Sci.*, **54**, 5—23.

Sardeshmukh, P. D., M. Newman, and C. R. Winkler, 1999: Dynamically consistent estimates of diabatic heating. Proceedings, *24th Annual Climate Diagnostics and Prediction Workshop*, Tucson, AZ, 172--175.

Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate,* **13**, 4268-4286.

Schemm, J. E., H. M. van den Dool, and S. Saha, 1996: A multi-year DERF experiment at NCEP. Preprints, *11th Conference on Numerical Weather Prediction*, Norfolk, VA, Amer. Meteor. Soc., 47—49.

Simmons, A. J., 1982: The forcing of stationary wave motion by tropical diabatic heating. *Quart. J. Roy. Meteor. Soc.*, **108**, 503—534.

Simmons, A. J., J. M. Wallace, and G. W. Branstator, 1983: Barotropic wave propagation and instability, and atmospheric teleconnection patterns. *J. Atmos. Sci.*, **40**, 1363—1392.

Smith, C. A., and P. D. Sardeshmukh, 2000: The effect of ENSO on the intraseasonal variance of surface temperatures in winter. *Int. J. Climatol.*, **20**, 1543-1557.

Ting, M., and P. D. Sardeshmukh, 1993: Factors determining the extratropical response to equatorial diabatic heating anomalies. *J. Atmos. Sci.*, **50**, 907-918.

Van den Dool, H. M., 1983: A possible explanation of the observed persistence of monthly mean circulation anomalies. *Mon. Wea. Rev.*, **111**, 539-544.

von Storch, H., and J.-S. Xu, 1990: Principal oscillation pattern analysis of the 30- to 60-day oscillation in the tropical troposphere. Part II: Definition of an index and its prediction. *Clim. Dyn.*, **4**, 175—190.

von Storch, H., and D. P. Baumhefner, 1991: Principal oscillation pattern analysis of the tropical 30-60 day oscillation. Part II: The prediction of equatorial velocity potential and its skill. *Clim. Dyn.*, **6**, 1—12.

Whitaker, J. S., and P. D. Sardeshmukh, 1998: A linear theory of extratropical synoptic eddy statistics. *J. Atmos. Sci.*, **55**, 237-258.

Xie, P. and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539-2557.

Xu, J.-S., and H. von Storch, 1990: Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate*, **3**, 1316—1329.

Zhang, C., and H. H. Hendon, 1997: Propagating and standing components of the intraseasonal oscillation in tropical convection. *J. Atmos. Sci.*, **54**, 741—752.